EUnetHTA 21 Public Consultation Comments and Responses
Of D4.4 - Endpoints

International Association of Mutual Benefit Societies (AIM), Belgium
AstraZeneca (AZ), Europe
Bundesarbeitsgemeinschaft Selbsthilfe von Menschen mit Behinderung, chronischer Erkrankung und ihren Angehörigen e.V. (BAG SELBSTHILFE), Germany
German Medicines Manufacturer´s Association (BAH), Germany
European Association of Hospital Pharmacists (EAHP), Belgium
Edwards Lifesciences, Europe
European Federation of Pharmaceutical Industries and Associations (EFPIA), Belgium
European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) HTA SIG, Europe
European Organisation for Research and Treatment of Cancer (EORTC), Belgium
European Huntington Association (EHA), Belgium
European Confederation of Pharmaceutical Entrepreneurs (EUCOPE), Belgium
GKV-Spitzenverband – GKV-SV, Germany
GSK, Europe
IGES Institut GmbH and HealthEcon AG (IGES LifeScience), Germany
OAK Access, Netherlands
Les Entreprises du Médicament (Leem), France
SKC Beratungsgesellschaft mbH (SKC), Germany
European Union of General Practitioners/ Family Physicians (UEMO), Belgium
Verband Forschender Arzneimittelhersteller (vfa) e.V., Germany
HTAi Patient and Citizen Involvement in HTA Interest Group (PCIG), Global
European Society for Medical Oncology (ESMO)
Lymphoma Coalition, Lymphoma Coalition Europe (LCE), France
Medtronic (Mdt), Switzerland
Patient Focused Medicines Development (PFMD), Belgium
Quality HTA, Canada
F. Hoffmann La Roche (Roche), Switzerland
European Hematology Association (EHA), the Netherlands
ISPOR – The Professional Society for Health Economics and Outcomes Research
Bayer AG & Bayer Vital GmbH, Germany
Takeda Pharmaceuticals International AG (Takeda), Belgium, Switzerland
MedTech Europe (MTE), Belgium
European Organisation for Rare Diseases (Eurordis), France
European Alliance for Vision Research and Ophthalmology (EU EYE), Belgium

EUnetHTA 21 Public Consultation Comments and Responses
Of D4.4 - Endpoints

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
James Ryan AstraZeneca	NA	All	Thank you for the opportunity to respond to the consultation. AstraZeneca have responded to the consultation via EFPIA, whose response represents the considerable knowledge and expertise of HTA, clinical, safety, COA, PRO, and HEOR experts across companies. We fully support the EFPIA responses and look forward to seeing your careful consideration of them in the preparation of the final document.	Thank you!
EAHP	General	-	The joint clinical assessment reports foreseen in point 23 of the health technology assessment regulation' preamble and that this Individual Practical Guideline document touches upon, should be shared with healthcare professionals in a transparent and timely manner, to ensure that they have all the information about a product when considering it for the patient's treatment.	Thank you for this comment.
Mihai Rotaru, EFPIA	General		**1. Consideration of external guidelines in setting endpoints** The draft guideline does not currently adequately reflect, nor mention, important guidelines or documentation following international standards of evidence-based medicine. This includes those by several Regulatory Agencies on the topic of selecting and developing *fit-for-purpose* clinical outcome assessments (including PROs), an important aspect of the broad topic of patient-focused drug development. In addition, EFPIA recommends that the guidance explicitly references the ICH E9(R1) Addendum, given its increasing importance for the design, conduct and analysis of clinical studies. The ICH E9(R1) Addendum specifies the estimands attributes, including a description of a "variable or endpoint", defined as: "The variable (or endpoint) to be obtained for each patient that is required to address the clinical question. The specification of the variable might include whether the patient experiences an intercurrent event". Link to the 2019 ICH E9(R1) guideline: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf	This guideline is intended to provide practical guidance for allowing HTAb to perform JCAs which takes place after studies are conducted. It is not intended to be a guideline on outcome development. "Fit for purpose" COA is a concept close to validity, reliability and interpretability as described in section 5.However, we will consider clarifications on how these concept overlap in the respective section of the guideline. Regarding the estimand framework which can be considered as an extension of the PICO approach, we will clarify overlap in the next version of the draft. Future collaboration between EMA and EUnetHTA 21 are out of scope of this guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Moreover, it would be useful to reference the EMA-EUnetHTA21 Bilateral Meeting minutes as the European Commission has pointed out the importance of the collaboration between HTA bodies and EMA, especially to build the new system under the HTA Regulation. Link to the June 2022 Bilateral Meeting minutes: https://www.eunethta.eu/wp-content/uploads/2022/10/EUnetHTA21-EMA-Bilateral-Meeeting-june-2022_en.pdf	
Mihai Rotaru, EFPIA	General		**2. Consideration of regulatory agencies** EFPIA recognises that the identification of the "outcomes" relevant for a given JCA is expected to reflect the (policy) question of interest across Member States, expressed through the relevant PICOs in the scoping process. However, the D4.4 guideline should explicitly acknowledge that, at the time the PICO definition is expected to take place, the relevant clinical trial(s) informing the JCA can be expected to have already been completed or had its primary readout. Since the design, conduct and analysis of clinical trials is also informed by the expectation of Regulatory Authorities, understanding their requirements on the topic of "endpoints" is particularly important to ensure that a given clinical trial is able to meet different stakeholders' needs. In particular, EFPIA recommends that the D4.4 Guideline references key guidance that the European Medicines Agency (EMA) has issued on the clinical evaluation of anti-cancer therapies and other disease areas, which contain important provisions on the topic of endpoints, such as: 1. The 05 January 2019 EMA/CHMP/205/95 Rev.6 Committee for Medicinal Products for Human Use (CHMP) Guideline on the clinical evaluation of anticancer medicinal products, which states that clinical endpoints such as OS, PFS, EFS, and DFS "can all be considered as adequate primary endpoints in confirmatory clinical trials to measure clinical benefit" https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-evaluation-anticancer-medicinal-products-man-revision-6_en.pdf	According to the HTAR, outcomes are requested during the assessment scope based on MS needs (Article 8(6)), which do not have to provide a rationale for these requests.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			2. The 13 December 2012 EMA/CHMP/27994/2008/Rev.1 Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man, covering methodological consideration for using progression-free survival (PFS) or disease-free survival (DFS) in confirmatory trials (https://www.ema.europa.eu/en/documents/scientific-guideline/appendix-1-guideline-evaluation-anticancer-medicinal-products-man-methodological-consideration-using_en.pdf) 3. The EMA/CHMP/292464/2014 Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man, on the use of patient-reported outcome (PRO) measures in oncology studies (https://www.ema.europa.eu/en/documents/other/appendix-2-guideline-evaluation-anticancer-medicinal-products-man_en.pdf) 4. The EMA - Clinical efficacy and safety guidelines (https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines/clinical-efficacy-safety-guidelines)	
Mihai Rotaru, EFPIA	General		**3. Outcomes relevant for HTA are not just final or patient-centred endpoints; the totality of evidence should be considered including clinical endpoints** Patient-centred endpoints, including impact on overall survival, symptoms and HRQoL are an important endpoint for many decision-makers, not just HTA bodies, and we support their inclusion in the joint HTA. However, EFPIA's experience across multiple HTA bodies in the EU and around the world show that these are not the only endpoints of interest, and we therefore disagree that only patient-centred endpoints, or surrogates for these endpoints, are the only relevant ones for EU HTA. Clinical endpoints, including primary and key secondary outcomes, are not only used as a surrogate for patient-centred outcomes in clinical trials, where many trials also collect patient-centred outcomes (as defined by EUnetHTA21) as part of their outcome set. They are used to establish clinical benefit, an important consideration	Outcomes are requested by MS as they see fit. As a general guidance, we think in the context of HTA assessment, patient-centred outcomes are appropriate for assessing the added value of a technology, but we do not restrict what MS members can ask. JCA are driven by PICO questions. PICO (assessment scope) will be performed blinded and will not be data driven (see guideline 4.2 Scoping process). There is no rationale to systematically include all outcomes included in the trial.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			in JCA and across HTA bodies. PROs are also intended to supplement efficacy and safety data with patient-generated data. These data inform a more comprehensive understanding of the clinical benefit of treatment by providing rigorous evidence of the treatment's impact on functioning, quality of life or symptoms evolution overtime. Specifically, clinical endpoints are relevant endpoints in their own right, and are important for HTA purposes: 1. Clinical endpoints and their magnitude of benefit are important. Even in the absence of evidence of surrogacy to patient-centred outcomes at time of approval, a large magnitude of benefit in a primary clinical outcome is highly relevant for JCA. Importantly, this can only be determined during a JCA itself and so it would be detrimental to not include a clinical outcome upfront at scoping, particularly if the clinical benefit is of a sufficient magnitude that regulators, clinicians and patients themselves would see this as transformative of patient care. 2. We recognise that a decision at a Member State should not be based on a single clinical endpoint in isolation to other endpoints. An improvement in a clinical endpoint also needs to show consistency across other endpoints and therefore relevant secondary endpoints are important to include. For example, an improvement in response rate in a disease is more important if there is also a prolongation of disease without progression. Likewise, an increase in PFS that is also supported by an increase in PFS2, should be a strong indication that the patient's disease has benefitted from the intervention. 3. The role of HTA differs across Member States, and can support clinical practice and guidelines, as well as health care delivery. Clinical practice is often based, in conjunction with patient-centred considerations, on clinical endpoints including treatment decision choice and when to stop a treatment and switch to another. 4. Clinical endpoints can provide reassurance to patient-centred endpoints themselves – overall survival can be subject to confounding bias (i.e. due to treatment switching) and patient reported outcomes from clinical trials can have missing data. In both cases, a clinical endpoint with a sound clinical and biological rationale can provide additional information and insight on the	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			interpretation of the patient-centred outcome, even though it may not show surrogacy. In clinical practice itself, a clinician may rely on both physiological markers (e.g. tumour size) and symptom reporting by the patient to make a judgement on treatment. 5. The disease context can also be critical. For example, in the field of vaccines, immunogenicity is a very specific clinically relevant outcome for estimating vaccine efficacy that is widely used in clinical trials and required when direct efficacy studies are not possible to conduct. WHO, and EMA guidelines (1,2) provide details on how these should be used and developed in clinical trials. Immune correlate of protection (CoP) is an immunological assay (either humoral or cellular immune response) that reliably predicts protection against disease or infection after vaccination or natural infection. A CoP is of great importance because it can be used as a surrogate endpoint assessing vaccine efficacy without directly observing clinical endpoints. The availability of this kind of surrogate endpoints help to avoid large-scale field efficacy trials and facilitate getting vaccine candidates approved. However, even when there is a validated correlate of protection, it is not always possible to translate a level of immunogenicity into a final efficacy outcome. 6. Economic assessment – many HTA systems also use cost-effectiveness analysis as part of their methodology. Clinical endpoints can form a critical part of the analysis, defining disease severity, health states, treatment related endpoints that are used in both the health benefit and cost part of the economic models. Although the JCA excludes economic assessment, the outcomes from the JCA can provide input to subsequent national economic analyses. 7. Consistency and transparency – trust in a HTA system is multi-factorial and consistency and transparency in decision-making are important components of this. For society and patients, there needs to be consistency not just in the use of clinical endpoints across decisions in a disease area, but importantly where they are used in patient-clinician decision-making and where there is clinical consensus on their utility. In summary, our proposal is that clinical endpoints included in the clinical trial, and used for regulatory decision-making, clinical guidelines, and clinician-patient interactions should be considered in the JCA, where their relative benefit can be	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			assessed. Excluding these critical endpoints, or at best restricting to those that have shown surrogacy to only patient-centred and final endpoints, falls short of good evidence-based medicine practice, undermines the intent and scientific credibility of JCA in delivering to the needs of many Member States, and will lead to sub-optimal appraisal decisions at a Member State level. References: 1. WHO Technical Report Series 1004, Annex 9, 2017. Guidelines on clinical evaluation of vaccines: regulatory expectations. 2. EMA 2006 Guideline on clinical evaluation of vaccines (EMEA/CHMP/VWP/164653/2005)	
Mihai Rotaru, EFPIA	General		**4. Inconsistency between Scoping proposal and Endpoint guideline** EFPIA notes that there is inconsistency between this guideline's objective that is "to provide guidance for MS in defining relevant outcomes during the scoping process" for use in the JCA compared to the Scoping proposal that states "It is the responsibility of the MS to define the PICO parameters according to their national legal and procedural requirements." As outlined above, EFPIA strongly believe that the key clinical endpoints, patient-centred outcomes and final endpoints are important outcomes for HTA bodies and Member States through the EU, and guidance around their applicability at a scoping stage contradicts the principle of Joint Clinical Assessment, international standards of evidence-based medicine, and the needs of many HTA bodies. EFPIA continues to call for a set of harmonised European PICOs based on a transparent, evidence-based process giving due consideration to clinically relevant populations and comparators restricted to the best available. Populations, comparators and specific outcomes, such as safety, requested by a minority of Member States should be handled with the use of complementary analyses, as enabled by the Regulation.	The purpose of this guideline is to propose some guidance for helping the MS to specify their needs in terms of outcomes. It is not in the contradiction with the fact that in the end, it is the responsibility of MS to define PICO questions as they see fit. The scoping process is not the purpose of this guideline.
Mihai Rotaru, EFPIA	General		**5. Request to update the five existing EUnetHTA endpoint guidelines** EFPIA would like to thank EUnetHTA21 for the opportunity to provide comment and	The purpose of the EUnethTA 21 contract was to develop a guideline according to a pre-specified project plan. Future works,

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			suggestions to improve the applicability of D.4.4 Outcomes (Endpoints) for European JCA. We note that EUnetHTA21 Project plan for D4.4 Endpoints[1] outlined a deliverable that we feel does not appear to have been addressed in the existing draft document. In particular, EUnetHTA21 outlined they were to check the existing EUnetHTA endpoint guidelines[2-6] for consistency and to consider updating them. We appreciate that time constraints may have prevented the authors of this guideline from actioning these two deliverables. If this is the case, we propose that an additional section "Future Recommendations" is added to the end of the draft guideline. Proposed wording for this section could read "*Given that new clinical outcomes may have to be designed for novel cancer therapies and ATMPs that will undergo EU JCA from 2025, and for other therapy areas in the future, we recommend that the existing EUnetHTA endpoint guidelines[2-6] are reviewed under the direction of the future Methodological Subgroup every 3 years.*" The time interval for the endpoints guidelines reviews would be consistent with that proposed in D.4.3.1 Direct and Indirect Comparison[7]. The five existing EUnetHTA endpoint guidelines have been published in 2015 and EFPIA believes that the guidelines should be based on based on state of art techniques in the field of health outcomes and clinical research, and, as outlined in the Regulation, "international standards of evidence-based medicine." The approach should evolve over time reflecting new analytical challenges emerging from the evolving medicine development paradigm and academic developments in the scientific areas related to HTA. We would like also to underline that the guidelines related to "Health-Related Quality of Life and Utility Measures" discussed economic evaluation extensively which is not relevant for the purpose of the EU JCA. References: 1. European Network for Health Technology Assessment EUnetHTA 21 Project Plan D4.4 Endpoints. Version 1.0 03/12/2021	such as updating previous guidelines, will be decided after the constitution of the coordination group under the HTAR.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			2. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Clinical Endpoints. Adapted version 2015 3. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Safety. Adapted version 2015 4. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Surrogate Endpoints. Adapted version 2015 5. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Health-Related Quality of Life and Utility Measures. Adapted version 2015 6. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Composite Endpoints. Adapted version 2015	
Mihai Rotaru, EFPIA	General		**6. Value of dialogue with health technology developers (HTDs) in scoping** EFPIA strongly believes that any direct interaction between the EU HTA Coordination Group and the HTD before the finalization of the scoping process would be useful to ensure an appropriate selection of the relevant clinically relevant endpoints, which form one of the critical elements of the PICO. As highlighted in EFPIA's comments during the public consultation about the scoping process, there should be adequate touchpoints at the right time in the JCA process, particularly through a well-designed scoping process, for the respective manufacturer to provide their input and elaborate on the evidence submitted, respond to requests but also to reflect on the PICO. Involvement of the HTD in the scoping process (and beyond) can help provide further insight to the evidence available (including timepoint assessments) which would make the process more efficient and relevant, allowing the assessors to consolidate the PICOs further when multiple outcome measures are requested for an outcome. Furthermore, an interactive PICO meeting, with all key stakeholders present, including the HTD, would be beneficial to the quality, efficiency, and representativeness of the JCA process. This would ensure that the HTD can provide a comprehensive and complete dossier based on available data which is aligned to the expectations of the JCA assessors and increase the scientific credibility of the final JCA report itself.	Interactions between HTDs and HTA bodies during the whole process of JCAs and JSCs are out of the scope of this guideline.

EUnetHTA 21 Public Consultation Comments and Responses
Of D4.4 - Endpoints

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Mihai Rotaru, EFPIA	General		**7. Inclusion of clinical experts in selecting appropriate endpoints** The choice of relevant outcomes should be based on international standards of evidence-based medicine. The input from healthcare professionals and clinicians should be sought in order to ascertain relevance of outcomes and to provide context specific considerations. EFPIA suggests to clearly specify in this guidance document how and when these inputs will be collected for the selection of appropriate endpoints. EFPIA also recommend introducing explicit wording that context specific considerations may be needed to select the appropriate endpoints.	Inclusion of stakeholders in the whole process of JCAs and JSCs are defined in other guidelines.
Mihai Rotaru, EFPIA	General		**8. Inclusion of patient experts in selecting appropriate endpoints** We recognize that ensuring COA (clinical outcomes assessment) and other measures evaluated in JCA are both patient-centred and patient-developed represents methodological and practical challenges[1]. The patient perspective is not only relevant for PROM selection, but also generally they need to have a say in the choice of outcomes. Other COA may be equally relevant from the patient perspective, such as PerfO, ObsRO and even ClinRO (for example in mental health). We believe that JCA must include knowledgeable patient experts in this evolving discussion and ask this language be included in the final guidelines: "Patient experts should be included in the review of patient-centred outcomes and in the selection of other relevant data for their unique and valuable perspectives." References: 1. Selby P, Velikova G. Taking patient reported outcomes centre stage in cancer research - why has it taken so long? Res Involv Engagem. 2018 Jul 19;4:25. doi: 10.1186/s40900-018-0109-z. PMID: 30038798; PMCID: PMC6052546.	Inclusions of patients in the whole process of JCAs and JSCs are defined in other guidelines.
Mihai Rotaru, EFPIA	General		**9. Guidance on minimally important difference (MID)** EFPIA welcomes the inclusion in the practical guideline on the interpretability of scales and on minimally important difference (MID) in particular. At the same time, EFPIA wishes to stress the importance that evidence based principles inform the generation and use of MIDs.	We agree MIDs can vary according to various characteristics such as baseline state etc. However, there is a balance to be found between providing sufficient guidance to assessors and not writing the guideline as a statistical textbook. We will consider if minor expansions are needed

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Furthermore, we encourage the guideline to further reflect that MIDs on the same scale can vary according to the therapy area, disease severity and line of treatment. More specifically: • The MID can depend on where a patient's score lies on the scale (e.g. early-/late-stage disease, due to non-linear progression along scale or non-linear association with health-related quality-of-life[1-5]) and the direction of change • The particular disease and patient population (if a generic measure is used, the MID might be different in different populations) • The duration of the trial (MID may need longer to meet than trial duration) MIDs should always reflect meaningful benefit to the patient. Purely distributional methods which focus on the minimal detectable change or standard error of measurement can be misleading and should be validated by the patient and clinical perspective the patient and clinical perspective[6]. For that reason, the guidance on distributional methods should be revised to encourage the use of external anchors or HRQoL instruments[7] and a flexible approach to validated thresholds. We welcome additional language to strengthen MID's being considered in the context of the JCA. Finally, we request the guideline reflect that both absolute and relative changes can be used for MIDs. **References** 1. https://doi.org/10.1111%2Fene.13941 2. https://doi.org/10.1179%2Fjmt.2008.16.4.82E 3. https://doi.org/10.1186/s12955-020-01344-w 4. https://doi.org/10.1186/1477-7525-12-66 5. https://doi.org/10.1016/j.trci.2019.06.005 6. https://doi.org/10.1007/s11136-004-0705-2 7. https://doi.org/10.1097/00005650-199905000-00006	for the next version of the draft. We also agree distribution-based methods do not consider patients' perspective. It is already stated within the guideline that anchor-based methods are considered more appropriate because they can take into account this perspective. Nonetheless, as distribution-based methods MIDs are still prevalent, there was a need to briefly describe these kinds of MIDs for helping assessors in identifying them during the process of conducting a JCA.
Mihai	General	Safety	**10. Safety endpoints and risk of misleading results**	These considerations will be considered for the next version of the draft.

EUnetHTA 21 Public Consultation Comments and Responses
Of D4.4 - Endpoints

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Rotaru, EFPIA			EFPIA is concerned that the draft proposal could lead to a substantial number of safety analyses that would be substantially disproportionate to the purposes of the JCA. In particular, allowing Member States to request additional safety analyses beyond those outlined, is likely to lead to many more exploratory safety requests, particularly from Germany who have a unique need for these exploratory analyses. Recent publications have shown that the safety section of IQWiG submissions averages nearly 600 analyses across an average of 1.2 PICOs (vfa, 2021). Oddens et al. (2022) reported in their case study "The safety-related exploratory analysis of verubecestat led to 206 statistical analyses for treatments and 812 treatment-by-subgroup interaction tests." Recent analysis by EFPIA in a common NSCLC setting, following the Scoping proposal by EUnetHTA 21, has indicated that across the multiple PICOs identified, there would be thousands of requested safety analyses if German requests are incorporated. EFPIA wishes to stress that these analyses would rarely have been pre-specified and would mostly be post-hoc analyses. In addition to the questionable usefulness to most Member States in reviewing these analyses, such extensive analyses pose considerable Type I errors, or false signals (either positive or negative), increasing the risk of findings due to chance and compromising the interpretability of the analyses at a Member State level. As well as potentially contradicting the EMA, whose responsibility it is to establish the overall risk:benefit of a new technology, the publication of such analyses could lead to sub-optimal treatment decisions by those not aware of the issues due to these potentially unreliable analyses, for which the Coordination Group would need take accountability for. Therefore, EFPIA proposes that adverse events (AEs) evaluated in the JCA be focused on the very common (≥1/10) and common (≥1/100) as listed in the draft SmPC[1] plus any serious adverse events and any additional clinically relevant AEs as defined by the Regulator and clinical experts. These AEs are proportionate to the needs of JCA and reflect the most important aspects of safety which are relevant to clinical decisions between the clinician and patient when choosing a treatment. EFPIA proposes that no further safety requests are included in the EU JCA and, if more are needed, these are handled as complementary analyses at a Member State	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer

			level.
			
			In addition, EFPIA recommends that the EUnetHTA safety guideline[2] is updated in order to reflect the following points:
			
			• A clear description of the remit and scope of the future JCA relative safety assessment would be beneficial to avoid overlap or repetition with the regulatory safety assessment. Provide guidance on cross-functional collaboration and sharing of methodologies/best practices on safety assessment between the regulatory and HTA bodies
			• The guideline should consider including recommendations on:
			  • Systematic process and search strategies to identify all relevant safety data
			  • Process or methodology of the comparison of safety data between technology and comparators, in particular where observational or single arms studies are being utilised
			  • Consistency in safety reporting (i.e., who determines the severity of the AE (patient or physician))
			  • Clearer separation between initial and repeat safety assessment
			• Recommend considering all interventions (pharmaceutical, medical devices, and non-drug therapies) in comparative safety assessments and how to evaluate the relative safety of interventions with data from different groups
			• Advocate for guidance on the assessment of safety, with connected e-Health devices alongside medications, and how tools allowing reporting of AE severity by patients should be used
			
			Reference:
			1. European Medicines Agency. How to prepare and review a summary of product characteristics. https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/product-information/how-prepare-review-summary-product-characteristics
			2. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Safety. Adapted version 2015
			3. vfa, July 2021, http://www.vfa.de/report-amnog-dossier-requirements.pdf	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			4. Oddens BJ, Agaku IT, Snyder ES, et al. Exploratory analyses of clinical trial data used for health technology assessments: a retrospective evaluation, BMJ Open 2022;12:e058146. doi: 10.1136/bmjopen-2021-058146	
Mihai Rotaru, EFPIA	General		**11. A comprehensive and resourced system for Joint Scientific Consultation is needed** The Joint Scientific Consultation should provide an opportunity for HTDs to have initial discussions with Member States about outcomes and outcome measurements. This early consultation should be used to identify and flag areas of concern from Member States and allow HTDs to ensure there is appropriate level of validation of endpoints available. The appropriate timing of the JSC will be a critical factor in enabling HTD to reflect on the design, conduct and analysis of relevant trials and address the requirements with respect to endpoints.	These considerations are out of the scope of the guideline.
Mihai Rotaru, EFPIA	General		**12. Request for ATMP specific guidance on endpoints:** We note an additional appendix was included in the current guidance document for Oncology medicines as these will be included in the first phase of Joint Clinical Assessments (JCA). However, specific guidance has not been included for advanced therapy medicinal products (ATMPs), which are also a part of the first phase of JCA implementation, and such guidance would be also of interest for orphan indications /rare diseases. ATMPs are potentially transformative therapies which are often launched on the basis of single arm studies powered to a primary clinical endpoint that is predictive of long-term clinical benefit. However, the most robust surrogacy methods are typically impossible to apply with the evidence base generally available for ATMPs at the time of launch. These novel therapies have unique challenges in clinical trial design, maturity of data (as they are potentially transformative or sometimes curative treatments), and available existing evidence to validate surrogacy (new MOA and possible new clinical endpoints). As the HTA Regulation states its intention “to strengthen HTA	The appendix was only added to propose definitions of endpoints which are mostly related to survival analyses, as these outcomes are sufficiently consistent and standardized. ATMP are medicines for human use based on genes, tissues, or cells. The products differ in mode of action, are used for treatment of various diseases, and have no common outcome measures as the intended effect is specific for a given patient group. It seems impractical to list possible outcomes of ATMPs as it was done for oncology products.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			methodologies applicable to personalised medicine", of which many ATMPs would qualify as, EFPIA propose that specific guidance for endpoints (and other method guidance, where appropriate) for ATMPs is developed, in the context of challenges faced by these interventions.	
Mihai Rotaru, EFPIA	General		**13. Consistency across EUnetHTA 21 guidelines** EFPIA would like to remind the importance of looking at the definitions of outcomes with a **holistic approach and considering the requirements presented in the previous consultations 'direct and indirect comparison'.** EFPIA is very much looking forward to increase predictability from the future JCA process and would like to avoid any situation where a specific endpoint is asked for a new drug, but it is impossible to perform any direct or indirect comparison because this endpoint has not been used for the comparator. Outcomes that are measured across studies should have some uniformity for comparative effectiveness, otherwise data from most outcomes may not be presented in the context of the treatment pathway and other treatments.	We agree consistency in outcomes definition is in general a way forward an enhancement of evidence synthesis. Some sections of the document such as definition of patient-centred outcomes or recommendations regarding core outcomes set are in line with this comment. However, it is a much more general issue than this guideline which comes after HTD development plan.
EFSPI	general		EFSPI would welcome clarification on how the prior EUnetHTA guidelines should be considered in context of this new guidance. As the 2015 clinical and safety EUnetHTA guidelines have been referenced, should these be considered alongside this guideline? As the HRQoL and composites guidelines have not been referenced, is this guideline intended to replace these prior documents? If the previous guidelines of 2015 are still to be considered, We suggest adding a section that summarises the existing guidances (reference 1-5) as well as describing the plans to maintain these guidances to reflect the evolving methods. We recommend that the existing EUnetHTA endpoint guidelines 1-5 are reviewed under the direction of the future Methodological Subgroup every 3 years. The time interval for the endpoints guidelines reviews would be consistent with that proposed in D.4.3.1 Direct and Indirect Comparison (6).	This guideline must be interpreted in regards to its purposes which are described in the introduction of the document. Recommendations for MS and assessors are explicit in dedicated boxes. The possibility of updating previous guidelines will be decided after the constitution of the coordination group of the HTAR after the end of the EUnetHTA 21 guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			We would also welcome a regular assessment of which endpoints are specified in the scoping process vs. their use by Member States to make decision on a new technology. We do understand the need of multiple Populations and Comparators to answer MSs' needs. Nevertheless, endpoints that are asked for in the scoping process, but never used in practice, could be streamlined. One option would be to check whether all endpoints required are mentioned in national appraisals or to organise surveys with the MSs. 1. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Clinical Endpoints. Adapted version 2015 2. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Safety. Adapted version 2015 3. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Surrogate Endpoints. Adapted version 2015 4. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Health-Related Quality of Life and Utility Measures. Adapted version 2015 5. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Composite Endpoints. Adapted version 2015 6. EUnetHTA21 Individual Practical Guideline Document. D.4.3.1 Direct and Indirect Comparison. Version 2. 26.07.2022	
EFSPI	general		A Joint Scientific advice before the scoping process starts should provide an opportunity for HTDs to have initial discussions with MS about outcomes and outcome measurements. This early consultation should be used to identify and flag areas of concern from MS and allow HTDs to ensure there is appropriate level of validation of endpoints available. This early scoping discussions should provide an initial guidance to MS at the time of the PICO development and selection of Outcomes of interest.	Interactions between HTDs and HTAb during the JCA process are not within the scope of this guideline.
EFSPI	general		EFSPI would welcome guidance on whether it is appropriate to include outcomes/endpoints that were not primarily defined in the scoping process, provided that their use and importance is properly justified.	JCA aims at assessing evidence submitted by HTDs in order to answer PICO questions as defined by MS.

EUnetHTA 21 Public Consultation Comments and Responses
Of D4.4 - Endpoints

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
EFSPI	general		In general, and in particular in Section 2.1, EFSPI would welcome an alignment between the definition of 'outcome' in this guidance with the definitions in ICH E9(R1) and regulatory definitions. For example, under ICH E9(R1) a variable/endpoint is measured at the patient level and the variable/endpoint is one of 5 attributes of the estimand that together define the treatment effect of interest. Can you clarify how the definitions align with ICH E9(R1) and the estimand attributes, e.g. the effect measure is the summary measure under ICH E9(R1). A mapping of how the definitions of this guideline align to ICH E9(R1) would be helpful	This will be clarified in the next version of the draft.
EFSPI	general		It is suggested that the draft guideline includes guidance specifically about composite endpoints/outcome measures. The level of validity of a composite endpoint is often considered in relation to the types of individual components included, the direction of effect on the individual components, etc. (e.g. EMA Guideline on multiplicity issues in clinical trials). EFSPI would also welcome further guidance on newer type of endpoints such as digital endpoints and genetic markers.	We will consider if guidance on composite endpoints and digital endpoints is useful within the context of this guideline for the next version of the draft. Unless we have misunderstood the comment, genetic markers are usually used as diagnostic tools to select patients. In that sense, they are not outcomes as described in this guideline.
EFSPI	general		The Summary of Product Characteristics (SmPC) is the basis of information for healthcare professionals on how to use the medicinal product safely and effectively. Presently, it is not guaranteed that endpoints presented in the SmPC will be reflected in the publicly available JCA. From a patient and clinician perspective, such a situation will not contribute to the transparency of the evaluation of new medicines. Accordingly, EFSPI recommends that endpoints that are included in the Prescribing Information/ Summary of Product Characteristics be included in the JCA.	The main purpose of a JCA is an analysis of evidence submitted by HTDs in order to answer MS needs according to the PICO questions they have requested to draw their own conclusions at the national level. As JCA reports will be publicly available, we do not understand why it will not contribute to transparency. SmPC and JCA report will be complementary documents, and both can contribute to provide valuable info.
EORTC	General		We believe that there is a major missing starting point for the post approval questions. The regulator is expected to assure activity and therapeutic	General considerations about how clinical research should be conducted and

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			benefit. Most of the drugs do not come into a vacuum but add to an available therapeutic armentarium. Nevertheless, work does not stop but start at the time of marketing authorisation. Limiting the process to possibly a PICO exercise on regulatory dossier will not be patient centric but will remain drug centric. Questions which complement the regulatory dossier and deliver datasets which are needed for patients and health care systems need to be identified based on which relevant outcome and methodology will need to be decided. These questions may address the following non exhaustive list of issues: optimal dose, schedule, sequence, combination etc.. Numerous questions will be de-escalation of treatments, at least in oncology, these questions do not pertain to the commercial sector and are critical for society. It is of utmost importance that EUnetHTA revisit the start up of the process. The document does not bring convincing evidence that the ultimate user is central to this document in absence of a structural involvement of the field actors at the start.	interactions between HTDs and HTAb are out of the scope of this guideline.
EORTC	General		The relevance of the questions comes before the outcome can be decided. The scientific and methodological process which will identify the relevant questions is crucially missing	The process that leads to the generation of PICO questions is described in the scoping process guideline.
EORTC	General		The relevance and selection of the end-points cannot be dissociated from the design and this is missing on the document hence please refer to other general comments by the EORTC. For example PFS, a relative surrogate end-point in oncology may have a different meaning and impact if the design compares 2 treatments given at the same time, as opposed as if it addresses the sequence of the treatments (immediate vs delayed) this valuing the treatment depending on the evolution of the disease in a sequence question of treatments, reflecting the reality rather than a silo approach of a drug reaching the health care systems, as it seems the case in this document	JCA reports aims at factually assessing the evidence submitted by the HTD according to their PICO questions. JCA reports will contain elements describing designs and methods of the assessed evidence but will not conclude on the appraisal of the clinical added value of a health technology which will be left at national level. During this appraisal process, MS will have the liberty to ponder the relevance of results in regard to methodological elements the way they see fit.
European	general		MS have a lot of authority on decision making (with regards to what is considered a desirable outcome) and interpretation of the same data set	The HTAR does not call for such a harmonization of appraisal of the clinical

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Huntington Association			send for assessment. Consensus on a European level should be considered to allow for equal and fair approval of medication throughout the EU.	added benefit of a treatment.
Matias Olsen, EUCOPE	General		In general, the needs of Member States for assessment of innovative pharmaceuticals must be communicated as early as possible, and ideally well in advance of the start of a Joint Clinical Assessment, to allow developers to generate the required data during the clinical trial stage. Joint Scientific Consultations should therefore be offered to all developers.	Such considerations are out of the scope of this guideline.
Matias Olsen, EUCOPE	General		Overall, the guideline reads well and reports key elements for the assessment of clinical endpoints. However, compared to the previous EUnetHTA guidance on use of endpoints published in 2015, there are two key concepts that require further elaboration in the updated guidance: 1) The absence of data on final or long-term endpoints should be accepted when a clinical endpoint is difficult or impossible to study (very rare or delayed) or the target population is too small to obtain meaningful results on relevant clinical endpoints even after very long follow-up (very slowly progressive and/or rare diseases). 2) The possibility to demonstrate effectiveness on final clinical morbidity/mortality outcomes at a later stage following Marketing Authorisation approval, i.e. during reimbursement. In general, we would encourage stronger recognition that demonstrating a morbidity/mortality outcome within the timeframe of a clinical trial is unfeasible under certain conditions.	This guideline is intended for appropriate reporting of elements pertaining to outcomes by assessors when conducting a JCA, in accordance with PICO questions as requested by MS. Appraisal of these elements such as the acceptance of specific endpoints in certain conditions are left at the discretion of MS.
M. Ermisch – GKV-SV	General		The guidance should be supplemented with further explanation regarding the use of composite endpoints such as MACE. HTD need to be aware that composite endpoints often combine single components of different relevancy to patients. Thus, the combination of these needs to be justified within the dossier and results for the single components need to be presented in addition to the composite results.	Already addressed issue.
M. Ermisch –	General		While the joint clinical assessment report should be factual and should not contain any value judgement, ranking of health outcomes, conclusions on the overall benefit or clinical added value of the assessed health technology, it is	The requirements for reporting proposed in this guideline aim at allowing MS to appraise the elements that need in order

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
GKV-SV			imperative that the HTAR gives information on the validity of each endpoint presented. To allow for this, HTD should be advised to give extensive information on the operationalisation of each individual endpoint presented and on the specific procedure of measuring the endpoint. Based on this information, national decision making on the relevancy and the ranking of endpoints is enabled.	to draw their conclusions.
M. Ermisch – GKV-SV	General		In addition, we are of the opinion that for all endpoints that were defined as relevant outcomes in the PICO adequate statistical analyses should be presented. Descriptive analyses alone are generally insufficient.	The guideline D4.5 is about aspects such as hypothesis testing. This guideline does not mean that descriptive statistics should only be presented. In general, the appropriate conduct of analyses according to good clinical and statistical practices is of the responsibility of the HTD.
GSK	General	General	The guideline in its current version summarizes general concepts but still leaves large room for interpretation and specific application of concepts between the member states (MS). Based on scientific rationale there should be more harmonization between MS regarding the following aspects: • Clinical relevance • Acceptance of surrogate endpoints • Acceptance of Responder definitions	In order to comply with the HTAR, appraisal of such aspects is left at the discretion of MS.
GSK	General	General	Most of the guidance relies on broad references to good clinical practice or good statistical practice, which while meaningful to a certain extent (as some decisions have to be made on a case-to-case basis) does not advance the JCA's key purpose of ensuring sufficient certainty about the acceptance of the approaches used by the HTD and on the implementation of good scientific practice in the specific situation of interest. Therefore, the key parameters for implementing outcomes, surrogate validation, validity, reliability and responder definitions should be determined in advance by a close exchange between assessors and HTD – on the basis of state-of-the-art scientific methods. The outcome of this exchange should be binding in the	These considerations are out of the scope of this guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			sense that the agreed methods are accepted in the final assessment and by the individual Member States in their decision-making.	
GSK	General	General	The guideline should be open to 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.	As for any scientific document, it is not meant to be definitive.
Ingrid van der Eijk, OAK Access			**General remark regarding EUnetHTA assessments as of 2025** The added value of a methodologically consistent approach for population-based assessments is clear. However, currently more and more novel (personalized) medicines and medicines for (very) rare indications are being developed, that might not score well in population-based assessments due to obvious reasons. Although the assessment can be methodology-wise correctly performed, the consequent end result is expected to lead to access issues in various countries. Although that is not part of the methodological assessment phase since those decisions take place in the national reimbursement procedures following the assessment, the EUnetHTA assessment is supposed to be a major source for reimbursement decisions. Therefore, I strongly believe that for this type of innovations, some kind of individual patient-based assessment/evidence on patient level would be more appropriate, to enable access to this type of innovations. Therefore, I would like to appeal for an opening in this solid assessment methodology that provides for deviation for such valuable innovations.	Since this comment is a general comment regarding future JCAs as of 2025, we have not answered the comment in light of D4.4. However, the comment may be considered in the future by the HTAR coordination group.
Laurent Petit, Leem	General		Importance of dialogue : **Early advice** There is a need to provide an opportunity for industry for a dialogue within the newly established European HTA framework to discuss outcomes and outcomes measurements. Such dialogue would avoid many difficulties with endpoints deemed not acceptable, not able to be validated or not relevant for disease context. **Patient involvement** Patient associations may be given the opportunity to contribute to the	Interactions between HTD and HTAb, and between HTAb and patients are within the scope of other guidelines.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			dialogue as to which endpoints are relevant in each disease context.	
Laurent Petit, Leem	General		For the specific case of PFS, only quoted in the Appendix A, it would be helpful that the guidance document provide more information on cases in which PFS can be considered a clinically relevant and acceptable outcome, and on situations in which this would not be the case. Same comment applies to PFS, DFS, EFS, ORR and other endpoints. The guidance at hand seems to follow a German-like reasoning, which has been proven in time to leave little room for endpoints such as PFS. **Suggestion :** **There is a need to clearly state the acceptability of specific endpoints such as these.** **Such clarification has for example been made both by EMA[1] and FDA[2], which have stated that PFS can be an important outcome to measure clinically relevant patient benefit.** **Suggestion :** **Clarify if the EMA position below regarding surrogates such as PFS is the same one to be adopted interpreting the guidance at hand.** 1. European Medicines Agency. Guideline on the evaluation of anticancer medicinal products in man; 2017. Available from: https://www. ema.europa.eu/en/documents/scientific-guideline/guidelineevaluation-anticancer-medicinal-products-man-revision-5_en.pdf. 2. 4. Food and Drug Administration. Clinical trial endpoints for the approval of cancer drugs and biologics - guidance for industry; 2018. Available from: https://www.fda.gov/media/71195/download	MS do not have to provide a rationale when requesting outcomes during the PICO process. Therefore, we can assume that if a MS requests PFS, it will consider the results provided the way it sees fit. In addition, what is a “german-like reasoning”?
Dr Daniel Widmer UEMO	General		This document largely considers complexity, chronic diseases and multimorbidity, in particular with the core outcome set method (252-260) allowing a global vision involving several stakeholders (262). The concern for security also seems to us to be well documented (chap. 4). We also appreciate the desire to give meaning to numbers and scores (5.3. qualitative	Thank you!

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			meaning of a score, meaning of change for the patient). We find in this very technical document the values defended in family medicine.	
Sebastian Werner vfa	General		The draft guideline gives general principles for the definition and assessment of outcomes (clinical relevance, safety, validity, reliability, and interpretability of scales). However, the guideline is often unclear about which evidence is required to demonstrate clinical relevance, validity, reliability and a meaningful interpretion of the data. This leaves HTD with great uncertainty about the possible result of the JCA and about the required data submission. Thus, the guideline should give more clarifications regarding these aspects. This would greatly support the predictability of JCA.	We think a large consensus has been achieved within the framework of evidence-based medicine on what constitutes evidence with high certainty of results and what is not. Other guidelines such as the D4.5 and D4.6 provides further guidance on this matter. In addition, the HTAR also allows for JSCs. Therefore, wo do not think HTDs are left with "great uncertainty" about what they should produce in terms of evidence.
Sebastian Werner vfa	General		The draft guideline implements specific definitions of health outcomes which clearly favour "patient-centred outcomes" and penalize "intermediate and surrogate outcomes" (incl. ranking these health outcomes based on their "uncertainty") and potentially limiting the inclusion and acceptance of relevant health outcomes. However, the HTAR calls for a broad inclusion of *"health outcomes"* in joint clinical assessment of (Article 8 [6]), indicating that *all health outcomes* must be considered in the assessment, while ranking of health outcomes is prohibited (HTAR, Recital 28). The vfa is concerned that EUnetHTA's approach will potentially limit the inclusion and acceptance of clinically relevant outcomes of the regulatory approval (often deemed "intermediate and surrogate outcomes"). Such limitations implemented in the guideline are not in line with the provisions of the HTAR. The vfa strongly calls for deleting or softening the parts of the guideline that can potentially limit the inclusion and acceptance of clinical outcomes relevant for regulatory approval. *All* health outcomes should be considered in the joint clinical assessment. Further, the vfa recommends that outcomes accepted in pivotal studies as the basis for regulatory approval should be considered in JCA as valid clinical outcome measures. The fundamental principle of evidence-based medicine to use the best available evidence to inform health care decisions, should be applied.	We do not think that proposing as a general guidance that patient-centred outcome are outcomes that matters in the context of HTA (and in clinical research in general) is a bold claim. Our guideline generally depicts strengths and weaknesses of considering different types of outcomes in this context. Nonetheless, as MS can request the outcomes they see fit, our guideline does not imply surrogate outcomes will never be considered in the context of JCAs.
Sebasti	General		EUnetHTA21 should work towards a harmonized approach of outcome	Such harmonization is not the concern of

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
an Werner vfa			assessment in JCA that reflects Member States common needs, considers EMA requirements, scientific consensus, patients' perspective and the HTD's evidence situation. Determining a common outcome set for JCA based on shared interest for Member States is most desirable, as it ensures feasibility. Outcomes that are only relevant to one Member States should be requested at the national level as part of complementary national clinical analyses if these outcomes are numerous in extent. The request of extensive lists of specific adverse events (cf. Germany), which would drive large numbers of analyses in submission dossiers and JCA reports should be avoided. The vfa recommends determining the common outcome set for each JCA considering European medical-scientific guidelines and international scientific initiatives aiming at standardization of outcomes (e.g., COMET Initiative, OMERACT or SISAQOL-IMI, ect.) as both can facilitate harmonisation. Patients' perspective on these outcomes should be considered as part of the scoping process that should also involve the medical scientific societies. EMA's scientific recommendations for outcomes should be considered for more alignment of European HTA and regulatory approval. Finally, the HTD's clinical evidence situation should be considered to inform and align Member States requests on possible outcomes and meaningful time points.	this guideline nor the concern of the current HTAR.
Sebastian Werner vfa	General		The JCA decisions about the validity and reliability of measurement scales, and an adequate interpretation using responder definitions (incl. MID) should be based on scientific rationale that is harmonized across Member States. The draft guideline should provide more details on criteria for acceptability regarding validity and reliability and responder definitions (incl. MID). EUnetHTA21 and the CG should use its best endeavours to reach a consensus on acceptability criteria and work towards a harmonized approach.	Such harmonization is not the concern of this guideline nor the concern of the current HTAR.
Sebastian Werner vfa	General		A constructive methodological exchange between the assessors and the HTD is needed to ensure highest level of quality of the JCA. The vfa strongly calls for establishing a PICO meeting with the possibility of dialogue about methodological aspects of the dossier preparation. The scoping process should consider the HTD's clinical evidence situation to inform and align Member States requests on outcomes and meaningful time points.	This concern is out of the scope of this guideline.

EUnetHTA 21 Public Consultation Comments and Responses
Of D4.4 - Endpoints

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Sebastian Werner vfa	General		The draft guideline does not contain a link to Joint Scientific consultations (JSC). JSC about the relevant outcomes should be carried out in parallel with EMA to align on the rationale and use of outcomes. The consultations should involve medical scientific societies and patient organizations to consider their input on the relevance of outcomes from patients' and the clinicians' perspectives. It is important that these consultations ensure predictability for the HTD. Thus, that requested outcomes in these consultations should not be changed by Member States at a later stage but should ensure consistency in the following scoping process.	JSCs are out of the scope of this guideline and have dedicated documents.
Sebastian Werner vfa	General		The draft guideline lacks information that addresses specific methodological approaches for new health technologies for which some data may not be readily available (such as ATMP, Orphans, ect.) as regulated by the HTAR (*"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." [Recital 24]).* The vfa recommends conforming with the HTAR and establishing adapted methodological approaches for joint clinical assessments and joint scientific consultation of ATMPs and Orphan Drugs, considering their specificities in their evidence generation with focus on health outcomes. Adapted methods for ATMP & Orphans should address specificities in all three domains of the guideline, regarding (i) clinical relevance, (ii) safety and (iii) validity, reliability, and interpretability of scales. Adapted methodological frameworks for the acceptability of intermediate and surrogate outcomes need to be established, as well as adapted surrogat validation frameworks that can appropriately cature the specificities of disease and technology.	As for the other practical guidelines on methodological and results elements, the aim is to allow factual reporting allowing MS to draw their own conclusions. Medical context is out of the scope of elements pertaining to certainty of results, including elements pertaining to outcomes.
Sebastian Werner vfa	General		The guideline lacks references to the previously developed EUnetHTA collaboration guidelines (e.g., composite endpoints, etc.) and does not provide an update of existing guidelines. The vfa recommends giving further clarification on composite endpoints, especially with focus on endpoint components. Further, the vfa would also appreciate guidance on new and advanced endpoints, such as digital endpoints.	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
HTAi PCIG	General		Despite the document clearly stating that (on LINE 262): "*By involving a wide range of stakeholders, such as patients, caregivers and health care professionals, it is more likely that patient-centred outcomes will be identified."* There is no patient involvement in the process outlined. The document offloads any patient involvement to pre-work by others to create Core Outcome Sets. But most diseases do not have core outcome sets, and will not have them for years, and by definition, Core Outcome Sets are a subset of the outcomes that could be measured in clinical studies. So, even when core outcome sets do exist, we would need patient perspectives on any additional outcomes and endpoints captured in the evidence. **Whilst COS are important, and should be considered, directly involving patients in the selection of outcomes is extremely important to ensure representation.** Core outcomes datasets are valid and useful tools in the selection of outcomes. Some of the limitations are highlighted within the guidance. Additional issues for some new and existing COS, which should be considered in their selection, are that they are often developed using multi-stakeholder Delphi surveys conducted in the English language and involving limited numbers of patients. Often the surveys are complex, so are completed by "expert" patients. Discussions around the PICO need patient input to give guidance of the patient-relevance. In terms of measures of outcomes and endpoints, <u>meaningful</u> patient involvement will be needed to ensure that the subsequent HTA activities are focused on the outcomes that matter to patients. (Meaningful patient involvement refers to patient involvement that enables informed dialogue (deliberation) for shared learning, problem solving and agreement). Patient involvement in the PICO discussion ensures the relevance of the PICO questions to actual patient need and ensures the alignment of the HTA to these needs. We are greatly concerned that the document as it stands fails to recognise the importance of patient input and involvement at this important stage of the HTA process.	We agree patients' involvement are of importance for defining relevant outcomes. In the context of JCAs, implication of patients is dealt within another guideline. Regarding how patients are considered when developing COS, this guideline is not the place to provide a methodological roadmap on how to develop a COS while allowing an appropriate account of the patient's perspective. Methodology for developing such outcome sets is evolving and we do hope that in the future patients' view will be better included. We will nevertheless consider if clarifications are needed for the next version of the draft.
European	General	n/a	ESMO believes that endpoints in Joint Clinical Assessments (JCAs) that are to be carried out under the EU's Health Technology Assessment (HTA)	We think the general guidance we have provided is in line with this comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Society for Medical Oncology (ESMO)			Regulation should be patient-centred endpoints, i.e. mortality and Patient-Reported Outcome Measures (PROMs) such as quality of life. Only in settings where such endpoints cannot be measured, surrogate endpoints can be used provided that there is surrogacy validation (Level I evidence: the treatment effect on the surrogate endpoint is similar or proportional to the treatment effect on the hard endpoint or Level II evidence: the association of the surrogate to a hard endpoint is confirmed). Additionally, we would like to highlight that the validation of surrogate endpoints should take into account disease settings and the intervention, as it may be tumour and drug- specific, and should be based on Level I evidence when possible.	
European Society for Medical Oncology (ESMO)	General	n/a	ESMO has developed the ESMO-Magnitude of Clinical Benefit Scale (ESMO-MCBS), which is a tool that uses a rational and structured approach to score the clinically meaningful benefit of medicines approved by the European Medicines Agency (EMA). The scale is used by various countries to prioritise cancer medicines and help frame the use of public and personal resources. ESMO considers that, for cancer medicines, the ESMO-Magnitude of Clinical Benefit Scale (ESMO-MCBS) may be instrumental in this process and help avoid possible duplication of efforts when used from the beginning when conducting JCAs. For further information, the contact details for the ESMO-MCBS Working Group are mcbs@esmo.org. Additional details on the ESMO-MCBS can be found here: https://www.esmo.org/guidelines/esmo-mcbs	Thank you for this information.
Marjorie Morrison, Lymphoma Coalition	General		**Endpoints for indolent and aggressive lymphomas.** With respect to endpoints in lymphoma, it is essential to consider the differences between indolent lymphomas and aggressive lymphomas. More specifically: Indolent lymphomas (or those that are the least aggressive) are characterized by their incurability. While they exhibit slow histology growth, new therapeutic interventions may prolong survival with	We thank you for this comment regarding this specific situation.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			some patients (such as those with follicular lymphoma) potentially living as long as the general population. As such, measuring overall survival after first-line therapy is an impractical study endpoint. Although overall survival and progression-free survival are clinical endpoints, patients with slow to progress lymphomas may require prolonged surveillance. There may also be influential factors that contribute to the variability in the response to initial treatments and outcomes that must be taken into consideration. In terms of the importance of clinical endpoints for patients with lymphoma, the overall survival endpoint is hampered by the length of time needed to assess. The progression-free survival endpoint (or progression itself) is not necessarily clinically relevant on its own given that a patient can live treatment-free with progressive disease for a longer period of time if their disease is slow-growing or asymptomatic – this points to the combination of different endpoints as being necessary to demonstrate treatment results accurately. Further, due to the time required to complete clinical trials and/or studies where overall survival endpoints and progression-free endpoints are applied, we perceive the risk of potential bias. This may occur with the introduction of subsequent therapies and/or potential delays in the development of newer therapeutic interventions that should also be considered (as treatment that addresses immediate disease control introduces potential longer-term consequences associated with prolonged therapy/risk of toxicities that patients may experience.) **Therefore, based on the differences between indolent and aggressive lymphomas, we wish to communicate the need for endpoints to take into consideration the additional needs of patients with lymphoma (such as, but not limited to, prolonged surveillance) that directly impact study**	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			endpoints.	
Marjorie Morrison, Lymphoma Coalition	General		**<u>Registries: Data Collection and Reporting.</u>** With respect to lymphoma, there remains insufficient data collection and reporting. This is driven in part by the lack of differentiating between the more than 80 lymphoma sub-types, and the lack of universal or consistent reporting requirements, practices and/or protocols. It is our view that endpoints that establish early or late progression of disease, complete response (and how long this can be maintained) and data that addresses quality-of-life is essential. However, with inconsistent data collection and reporting processes and/or practices, we question how this can be achieved. **We propose that fundamental data collection and reporting processes are applicable across EUnetHTA21 project and deliverables and until data is reported consistently and/or universally (essentially, universally reported health-related quality of life data and reporting in clinical trials) data are not accurately capturing or reflective the patient populations.**	We agree with such general comments, but it is not the goal of EUnetHTA 21 to address lack of data availability prior to JCAs in specific situations.
Natacha Bolanos, Lymphoma Coalition			References mentioned in the review: **References** (1) Lang KM, Harrison KL, Williamson PR, Huntly BJP, Ossenkoppele G, Geissler J, Bereczky T, Hernández-Rivas JM, Chevrou-Séverac H, Goodbody R, Schulze-Rath R, Bullinger L. Core outcome set measurement for future clinical trials in acute myeloid leukemia: the HARMONY study protocol using a multi-stakeholder consensus-based Delphi process and a final consensus meeting. Trials. 2020	Thank you.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			May 27;21(1):437. doi: 10.1186/s13063-020-04384-1. PMID: 32460828; PMCID: PMC7251906. (2) Pan-Stakeholder Core Outcome Set (COS) Definition for Selected Hematological Malignancies - Results of the Harmony Alliance. *Blood* (2021) 138 (Supplement 1): 5031.https://doi.org/10.1182/blood-2021-145002 (Accessed 24 October 2022) (3) Lang KM, Harrison KL, Williamson PR, Huntly BJP, Ossenkoppele G, Geissler J, Bereczky T, Hernández-Rivas JM, Chevrou-Séverac H, Goodbody R, Schulze-Rath R, Bullinger L. Core outcome set measurement for future clinical trials in acute myeloid leukemia: the HARMONY study protocol using a multi-stakeholder consensus-based Delphi process and a final consensus meeting. Trials. 2020 May 27;21(1):437. doi: 10.1186/s13063-020-04384-1. PMID: 32460828; PMCID: PMC7251906.	
Hayley Chapman, PFMD	General		There has been recent consideration of a taxonomy of impact, rather than outcomes (https://pubmed.ncbi.nlm.nih.gov/29288712/), to identify what is most meaningful and important to measure for patients. This has been applied in a recent initiative by the National Health Council where a blueprint for developing Core Impact Sets has been developed. Several organizations that are referenced in this document (including COMET (line 259) and OMERACT (line 255) were part of the NHC Steering Committee to help progress the thinking from outcomes to impacts. PFMD has also been working closely with the NHC, representatives from the patient community, industry and regulatory/HTA bodies to develop a navigator of Patient Experience Data, that references the need to prioritize the impacts that are most important to patients. It is suggested that consideration be given to patient-meaningful impacts, in addition to outcomes.	Thank you for this comment and for these references. However, at this stage, our guideline aims at providing guidance that is consistent with the scoping process which is conducted under a PICO framework.
Roche	General		The guideline does not sufficiently discuss the existing EUnetHTA methods guidelines developed previously (2015). We suggest adding a section that	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			summarises the existing guidelines (references 1-5) as well as describing the plans to maintain these guidelines to reflect the evolving methods. We recommend that the existing EUnetHTA endpoint guidelines 1-5 are reviewed under the direction of the future Methodological Subgroup every 3 years and under public consultation to enable engagement with HTDs. The time interval for the endpoints guidelines reviews would be consistent with that proposed in D.4.3.1 Direct and Indirect Comparison (6). 1. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Clinical Endpoints. Adapted version 2015 2. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Safety. Adapted version 2015 3. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Surrogate Endpoints. Adapted version 2015 4. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Health-Related Quality of Life and Utility Measures. Adapted version 2015 5. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Composite Endpoints. Adapted version 2015 6. EUnetHTA21 Individual Practical Guideline Document. D.4.3.1 Direct and Indirect Comparison. Version 2. 26.07.2022	
Roche	General		As mentioned in other consultations, we also have concerns about the lack of transparency. It is noticeable that the guideline lacks a certain depth and	Involvement of stakeholders in the process of JCAs and JSCs are the concerns of

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			details that leave many questions unanswered and could lead to unclear decisions being made. We understand that the practical guideline D4.4 cannot answer all questions. Therefore, we would like to reiterate how important and critical the involvement of the HTD can be in order to discuss such open questions fairly and accurately.	other guidelines.
Roche	General		Joint scientific consultation (JSC) is a crucial component of the EU HTA framework and provides an excellent basis for the discussion on the outcomes, outcome measurements, surrogate endpoints and validation requirements across the various stakeholders and decision makers (e.g. HTA, EMA, clinicians, patients, HTD). As such JSCs should be embedded into the guideline. The JSC should seek to establish alignment between HTA, EMA, and HTD on the acceptability of these outcomes, outcome measurements, surrogate endpoints and validation requirements.	This is out of scope of this guideline.
Roche	General		The individual PICO survey responses from all parties (e.g. HTA bodies, patients, clinical experts, etc.), including any divergent input, should be shared with the HTD along with the information about the final PICO(s) in a timely manner.	Thank you for this comment.
Roche	General		<u>Value of Stakeholder Engagement</u> Patient and healthcare professional involvement in relation to patient-centered outcomes is critical. This may include but is not limited to better understanding unmet need, outcomes that matter and the patient experience. Roche highlights that patient organisations and caregivers should be included in the early JSC discussions, the scoping process, as well as evaluation of outcomes during the JCA for a more comprehensive overview. These perspectives should be included throughout the process and within the guideline of the determination of outcomes.	Involvement of stakeholders in the process of JCAs and JSCs are the concerns of other guidelines.
Roche	General		<u>Inclusion of Totality of Evidence and Acknowledgment of Context Specific Considerations</u>	JCAs aims at providing the elements MS need to answer their PICO questions according to the scoping process. In general, we do not think elements

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			The JCA should evaluate all clinical outcomes defined by the HTD as part of the clinical study in order to provide a detailed and holistic assessment of the totality of clinical evidence submitted. Also, it is absolutely vital that assessors and co-assessors adopt a flexible approach to the outcomes and endpoints selected in a study. The set of outcomes/endpoints that is practically feasible may differ between therapeutic areas and/or types of intervention. The guideline must allow for such context specific considerations.This also applies to decision thresholds, such as meaningful change thresholds. Such thresholds should be established based on best scientific practice, which must be applied thoughtfully accounting for context specificities.	pertaining to factual methodological elements such as validity are context dependent. Regarding the appraisal of the magnitude of the benefit, it is ultimately left at the discretion of MS which can take into consideration the medical context.
Roche	General		Support for use of Composite Endpoints In the revised guideline, there is no mention of composite endpoints when these were specifically called out and had their own dedicated guideline in the 2015 versions. We feel that composite endpoints require guidance on their use and should be included in this guideline. Roche supports the use of composite endpoints where additional granularity is required in the endpoint, for example if a disease affects multiple organ systems, or if the single endpoint is not sensitive enough to accurately capture disease progression.From the previously developed guideline, further clarification is warranted in relation to the evidence needed to demonstrate clinical relevance. Roche would also like to highlight that qualitative patient or clinical input is fundamental for the selection of Clinical Outcome Assessments (COAs). Ref: European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Composite Endpoints. Adapted version 2015	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Roche	General		**10. Safety endpoints and risk of misleading results** Roche is concerned that the draft proposal could lead to a substantial number of safety analyses that would be substantially disproportionate to the purposes of the JCA. In particular, allowing Member States to request additional safety analyses beyond those outlined is likely to lead to many more exploratory safety requests, particularly from Germany who have a unique need for these exploratory analyses. Recent publications have shown that the safety section of IQWiG submissions averages nearly 600 analyses across an average of 1.2 PICOs (vfa, 2021). Oddens et al. (2022) reported in their case study "The safety-related exploratory analysis of verubecestat led to 206 statistical analyses for treatments and 812 treatment-by-subgroup interaction tests." Recent analysis by EFPIA in a common NSCLC setting, following the Scoping proposal by EUnetHTA 21, has indicated that across the multiple PICOs identified, there would be thousands of requested safety analyses if German requests are incorporated. Roche wishes to stress that these analyses would rarely have been pre-specified and would mostly be post-hoc analyses. In addition to the usefulness to most Member States in reviewing these analyses, such extensive analyses pose considerable Type I errors, or false signals (either positive or negative), increasing the risk of findings due to chance	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			and compromising the interpretability of the analyses at a Member State level. As well as potentially contradicting the EMA, whose responsibility it is to establish the overall risk:benefit of a new technology, the publication of such analyses could lead to sub-optimal treatment decisions by those not aware of the issues due to these potentially unreliable analyses, for which the Co-ordination Group would need take accountability for. Therefore, Roche proposes that adverse events (AEs) evaluated in the JCA be focused on the very common (≥1/10) and common (≥1/100) as listed in the draft SmPC[1] plus any serious adverse events and any additional clinically relevant AEs as defined by the Regulator and clinical experts. These AEs are proportionate to the needs of JCA and reflect the most important aspects of safety which are relevant to clinical decisions between the clinician and patient when choosing a treatment. EFPIA proposes that no further safety requests are included in the EU JCA and these are handled as complementary analyses at a Member State level. In addition, Roche recommends that the EUnetHTA safety guideline[2] is updated in order to reflect the following points:	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			- A clear description of the remit and scope of the future JCA relative safety assessment would be beneficial to avoid overlap or repetition with the regulatory safety assessment. Provide guidance on cross-functional collaboration and sharing of methodologies/best practices on safety assessment between the regulatory and HTA bodies - The guideline should consider including recommendations on: ● Systematic process and search strategies to identify all relevant safety data ● Process or methodology of the comparison of safety data between technology and comparators, in particular where observational or single arms studies are being utilised ● Consistency in safety reporting (i.e., who determines the severity of the AE (patient or physician)) ● Clearer separation between initial and repeat safety assessment - Recommend considering all interventions (pharmaceutical, medical devices, and non-drug therapies) in comparative safety assessments and how to evaluate the relative safety of interventions with data from different groups - Advocate for guidance on the assessment of safety, with connected e-Health devices alongside medications, and how tools allowing reporting of AE severity by patients should be used	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Reference: 1. European Medicines Agency. How to prepare and review a summary of product characteristics. https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/product-information/how-prepare-review-summary-product-characteristics 2. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Safety. Adapted version 2015 3. vfa, July 2021, http://www.vfa.de/report-amnog-dossier-requirements.pdf 4. Oddens BJ, Agaku IT, Snyder ES, et al. Exploratory analyses of clinical trial data used for health technology assessments: a retrospective evaluationBMJ Open 2022;12:e058146. doi: 10.1136/bmjopen-2021-058146	
EHA	General		Overall, we find the document balanced and well written.	Thank you!
Ioanna Psalti EUEYE	**General**		The EU EYE agrees with the ontology used in the EunetHTA document. It is the right way to go about the enormous task of identifying relevant outcomes.	Thank you for your comment. The guideline does not dismiss de facto any outcome measurement instrument, including disease specific QoL instruments, as long as the elements

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			The document distinguishes between clinically reported outcomes and technology assessed outcomes and recommends the use of "long term or final" outcomes for health technology assessment with mortality as the example. Intermediate or surrogate outcomes may be acceptable when it is not feasible to measure a final outcome,. For eye diseases the long term outcome translates to blindness or visual disability and a validated HRQoL measure that is sensitive to changes in improved or preserved vision or better joint function is required as opposed to death or severe physical handicap. It is difficult however to fully apply the guideline to diseases which remain asymptomatic until the late stages and linked with impairment/disability as a consequence such as glaucoma, diabetic retinopathy/maculopathy. Using a validated HRQoL to determine cost effectiveness seems to be the only way forward. It is not clear however to what extent scores of vision related QoL instruments like the NEI VFQ can be compared to other disease specific QoL measures such as those in rheumatology or cancer.	regarding their validity, reliability and interpretability can be assessed in the context of a JCA. We will consider if clarifications are needed for the next version of the draft.
EUEYE, Ioanna Psalti	**General**		**Expanding the use of the document** The EU EYE understands that it is beyond the scope of this paper to provide guidelines for all disciplines given that the first group of therapeutics to undergo JCA are oncology medicines. Understandably, the current EunetHTA document then reflects outcomes for assessing safety and efficacy of new cancer therapies and therefore Appendix A focuses in oncology outcomes. However the unintentional impact of this is a restriction of the usefulness of the document into diseases causing fatality or having early symptom onset. The use of the document can extent to diseases which remain asymptomatic until their late stages with disability as the final disease	Thank you for this comment. We will consider if it is useful provide some guidance regarding these issues in the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			progression rather than death if it were to include brief recommendations on how to adapt these guidelines for such diseases; or at least provide references to other EunetHTA work related to such diseases. Given that rheumatoid arthritis is mentioned in the document, it may be possible using it as an example of non-fatal disease to provide such brief recommendations as an additional appendix. Our comments may have further defined the problem rather than provide solutions and more than likely new EUnetHTA guidelines will follow before the HTAR enter its last phase in 2030. However simple recommendations to expand the use of the current document in the meantime will be of benefit given the recent findings on systemic comorbidity of chronic eye conditions and various types of cancer (Abedian, S., Abbasi, A., de Boer, A. *et al.* Effect of metabolic genetic variants on long-term disease comorbidity in patients with type 2 diabetes. *Sci Rep* **11**, 2794 (2021). https://doi.org/10.1038/s41598-021-82276-3; Lee SH, Ro JS, Chung KY, Lee SH, Park YL, Kim JE, Lee SH. Association between Skin Cancer and Systemic and Ocular Comorbidities in South Korea. J Clin Med. 2021 Jun 1;10(11):2451. doi: 10.3390/jcm10112451. PMID: 34205919; PMCID: PMC8198495)	
Ioanna Psalti EUEYE	**General**		**Endpoints** The forthcoming JCAs may include only a clinical assessment and cost-effectiveness assessment, value assessments and price negotiations or decisions remain in the domain of Member States. However, the decisions about the choice of the comparator and the endpoints will be made at the EU level which in turn will have an impact ob the price negotiations. The following points which are related to endpoints should therefore be considered either in this document or in future work:	Thank you for this comment. If indeed outcomes will be requested by MS during the PICO process, it does not prevent the fact HTAb will be attentive to the emergence of more relevant outcomes in specific areas.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			1. HTA has to consider how to allow for outcomes that move away from the conventional ones given that regulatory agencies begin to encourage drug developers to develop functional vision tests. Ophthalmology is moving from traditional/conventional to alternative alternative visual function endpoints, which are more sensitive to visual impairment than acuity. Conventional endpoints serve well when the disease is treated at a severe stage but it is difficult to reconcile many of these with the potential new drugs and outcomes when the vision is still good as in the cases of chronic sight-threatening eye diseases (glaucoma, diabetic retinopathy, maculopathy/) with visual loss occur in late disease stage. The functional endpoint may serve well orphan disease treatments, and precision medicine and new technologies, although they may not yet be able to provide approvable endpoints, they offer nevertheless more precise information on structural changes for proof of-concept endpoints in early clinical studies.	
Ioanna Psalti EUEYE	**General**		1. The document could provide additional guidance for overcoming the challenge presented by the paradigm shift triggered by the emergence of some novel endpoints in the marketing authorisation process for ATMPs. <u>Rationale:</u> The value of the connection between objective and subjective visual assessments is reinforced by the emerging importance of PROMs but some of the traditional performance tests are not appropriate for ATMPs and rare diseases. The	We will consider if additions are needed for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			example of Luxturna, the gene therapy for inherited retinal disease due to mutations, illustrates the issue well. Irrespective of whether the specific treatment is effective or not, the lesson from its HTA assessment is that data limitations associated with novel endpoints based on performance are problematic and that certain quality-of-life benefits may not be adequately captured in the HTA analysis as a result (Darrow JJ. Luxturna: FDA documents reveal the value of a costly gene therapy. Drug Discov Today. 2019 Apr;24(4):949-954. doi:Not 10.1016/j.drudis.2019.01.019. Epub 2019 Jan 31. PMID: 30711576).	
Ioanna Psalti EUEYE	**General**		1. A brief reference to the use of composite endpoints including a definition should be included (either in its entity or abbreviated) as it is present in EUnetHTA. 2015 Adapted version. Endpoints used for REA of pharmaceuticals: Composite endpoints" – February 2013; https://www.eunethta.eu/wp-content/uploads/2018/03/composite_endpoints.pdf *"A composite endpoint (CE) consists of two or more single events combined in one outcome that should represent an overall clinically relevant and valid measure of clinical benefit due to treatment. It is possible to combine binary or time-to-event endpoints. Either the occurrence of any event from a given set of events is of interest, or the time to the occurrence of the first event. Composite endpoints usually refer to combined morbidity and mortality endpoints. it may also be a combination of objective (e.g. laboratory measurements) and subjective outcomes (e.g. pain); in this*	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			case, clinical relevance of overall results can be more difficult to interpret." • Rationale: The importance of composite endpoints in multi-system diseases, cancer and rare diseases is increasing (McMenamin, M., Berglind, A. & Wason, J. Improving the analysis of composite endpoints in rare disease trials. Orphanet J Rare Dis **13**, 81 (2018). *https://doi.org/10.1186/s13023-018-0819-1),*	
Ioanna Psalti EUEYE	**General**		1. A table/appendix with the list of the emerging endpoints which are applicable to HTA evaluations will be useful. • Rationale: Emerging endpoints are identified as more relevant for newer products with improved efficacy and long-term survival and for multi-system diseases, cancer and rare diseases – including DFS, MRD, pCR, RR, and time to response. However not all of these may be applicable for HTA evaluations e.g. time to response.	We will consider if describing more endpoints in the appendix is necessary for the next version of the draft.
Ioanna Psalti EUEYE	**General**		The ophthalmology community recognises that defining common outcome measures even within one discipline is a complex endeavour - for example although acuity would be the most likely common outcome measure for ophthalmology, it may not be relevant to eye diseases/conditions such as glaucoma. The EU EYE believes that it would be of benefit if the document at least recognise such complexities and that the technical disease specific outcome measures upon which	The document is not intended to be a definitive list of outcomes MS can request, nor it implies that interdisciplinary outcomes should always be favoured. Nevertheless, we will consider if a clarification is needed for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			different disciplines rely do not necessarily convert into interdisciplinary outcome measures. This is particularly important for co-morbid conditions. The EU EYE https://www.eueye.org is an alliance of the major subspecialties in ophthalmology and has provided expertise for some of the pilot JCAs in the past. Our individual organisations work on achieving consensus on clinically meaningful outcomes for the various disease they represent. Our members are willing to work further with the EUnetHTA on the topic of endpoints and other related aspects of HTA.	
François Houÿez, Eurordis	general		There is no paragraph on Composite Endpoints, or scores that are distinct from Core Outcome Set defined in 3.2 Composite endpoints are used when the number of events (outcomes) is expected to be low, and for example it can combine number of patients who progress to a worse disease stage and/or death, when disease progression and death considered separately do not generate enough information to detect an effect. Even if composite endpoints are often difficult to interpret when none of its component taken separately provides statistically significant results, it can provide an indication for small population trials.	Already addressed issue.
MTE	General		Some further clarify will be of interest to define more clearly the remit. The Guidelines seem to limit the scope to clinical outcomes, while taken a patient centricity into consideration, ie mortality,morbidity,QOL and symptoms. However complementary outcomes that matter to patients , to health care professionals, to providers and health systems do not seem to be taken into consideration and it will be important to further clarify is the remit of the JCA is limited to clinical added value related information. Specific for medical technologies without a direct therapeutic effect but introduced within the care pathways, possible in combination with a therapeutic drug/device therapy we welcome further reflections to assess the	The main purpose of JCA is the appraisal of the relative effectiveness of a health technology in order to grade the clinical added value. Considerations that are described here are out of scope.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			intrinsic value.	
MTE	General		The guidelines seem to be lacking guidance related to combined endpoints, combining multiple outcomes	Already addressed issue.
MTE	General		With the QOL, we would welcome a more clear distinction and purpose of use between HRQOL and disease specific QOL	We will consider if such a distinction is needed for the next version of the draft.
Tanja Podkonjak, Takeda	general		**Totality of evidence and all clinically relevant outcomes should be considered for a JCA** Takeda would like to emphasise the importance of following the principles of evidence-based medicine in the context of EU HTA, which includes the consideration of the totality of available evidence during the JCA process. Takeda is concerned about the emphasis on only select endpoints of the available evidence, as proposed by the D4.4 Endpoints guideline, as it is arguably too narrow, focusing only on final or patient-centred outcomes (as defined) and would omit important evidence on the technology. Namely, clinically important endpoints (i.e., symptomatic response, progression-free survival (PFS), event-free or incident-free rates) are important outcome measures that inform clinical decision making, patient prognosis and health care resource utilisation (stop or modulation of treatment, inform future interventions). Clinical endpoints, including primary and key secondary outcomes, are not only used as a surrogate for patient-centred or final outcomes in clinical trials, but they are also used to establish clinical benefit, an important consideration in JCA, across HTA bodies and for clinical decision making. Taken in totality, clinical endpoints provide important information on the impact of the new intervention on the patient and health care system and omitting them from a JCA may impact the quality of the final report. In the field of oncology for instance, the guidelines should recognize that **PFS is both a patient-centred outcome and** "***an intermediate endpoint that is relevant on its own right",*** as stated in the EUnetHTA 2015 guideline on Clinical Endpoints.[1] A progression event impacts the course of current and	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			future treatments, provides information on patient prognosis (i.e. based on length of response, refractory status) and health resource utilization (i.e. frequency of visits, tests to be performed) – all important metrics of the effect the intervention or technology has had, providing valuable clinical information. A recently published study[2] found that although PFS as a clinical terminology may not be well understood by patients, time when the disease is not progressing is meaningful to patients, especially if it is coupled with improvement in HRQoL and no additional toxicity, and therefore is a patient-centered outcome. Clinical endpoints, patient-centred endpoints and final endpoints all contribute data to inform a more comprehensive understanding of the clinical benefit of new intervention by providing rigorous evidence of the treatments impact on functioning, quality of life or symptoms evolution overtime and should be considered in an EU JCA. References: 1. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Clinical Endpoints. Adapted version 2015. • Mertz S, Benjamin C, Girvalaki C, Cardone A, Gono P, May SG, Comerford E, Than KS, Birch K, Roach M, Myers S, Sasane M, Lavi L, Cameron A, Cardoso F. Progression-free survival and quality of life in metastatic breast cancer: The patient perspective. Breast. 2022 Oct;65:84-90. doi: 10.1016/j.breast.2022.07.006. Epub 2022 Jul 9. PMID: 35870420; PMCID: PMC9307669.	
Tanja Podkonjak, Takeda	general		**Inclusion of stakeholders and experts in selecting appropriate endpoints** We note the objective of the proposed guideline is to provide guidance for MS during the scoping process and to help assessors and co-assessors in	Interactions between HTAb and different stakeholders (HTDs, patients…) during the whole JCA and JSC processes are the concerns of other guidelines.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			conducting a JCA, however, we would like to highlight the absence of reference to stakeholder involvement in the selection of appropriate endpoints for the condition and technology in consideration. Takeda strongly believes that input from key stakeholders including patients, clinical experts and the HTD are instrumental in selecting appropriate and context-relevant endpoints or outcomes for a JCA. The choice of relevant outcomes should be based on international standards of evidence-based medicine. Input from clinical experts and patient experts, should be sought in determining how efficacy or outcomes are measured for the specific condition, what outcomes are important and how to interpret the results. In addition, we believe that direct interaction between the EU HTA Coordination Group and the manufacturer during the scoping process is important to the selection of relevant endpoints for the technology. The HTD has the best knowledge of the technology under assessment and the development plan and can therefore help provide further insight to the evidence available (including timepoint assessments) which would make the process more efficient and relevant. We would like to note that endpoints of a clinical trial and a technology's development plan are set many years in advance of a JCA scoping procedure, the JCA step the guideline aims to address, therefore input from the stakeholders involved in the development plan – which includes clinicians, patients, regulators, and the company - is instrumental in understanding the technology and selecting appropriate endpoints for a JCA. We reiterate our recommendation to establish an interactive PICO meeting in the scoping stage of a JCA, with all key stakeholders present, including the HTD, as a critical steps beneficial to the quality, efficiency, and representativeness of the JCA process. Beyond the scoping meeting, Takeda requests specific text and recommendations be added to the guideline D4.4 Endpoints on the role of the HTD, patient and clinical experts in setting appropriate endpoints for a JCA.	
Tanja Podkon	general		Takeda notes that composite outcomes are not mentioned in the current	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
jak, Takeda			guideline, and we request these be included in the guideline as potential outcomes. Furthermore, we request additional guidance be included on the use of digital endpoints for JCA either in this guideline or a planned future guideline.	
Roche	General	Section 2.1	Patient-Reported Outcomes (PRO), Observer-Reported Outcomes (ObsRO), Clinician-Reported Outcomes (ClinRO) & Performance Outcomes (PerfO) are all types of Clinical Outcome Assessments (COA) that measure how patients feel, function or survive (Walton et al. 2015). Roche proposes that the term COA is introduced (e.g., in line 77 or line 147) and used more systematically in this guideline. *Reference: Walton MK, Powers JH 3rd, Hobart J, Patrick D, Marquis P, Vamvakas S, Isaac M, Molsen E, Cano S, Burke LB; International Society for Pharmacoeconomics and Outcomes Research Task Force for Clinical Outcomes Assessment. Clinical Outcome Assessments: Conceptual Foundation-Report of the ISPOR Clinical Outcomes Assessment - Emerging Good Practices for Outcomes Research Task Force. Value Health. 2015 Sep;18(6):741-52. doi: 10.1016/j.jval.2015.08.006. Epub 2015 Aug 24. PMID: 26409600; PMCID: PMC4610138.*	Thank you for this comment. Of note, COA is the first word of the introduction of the guideline. But we will consider clarifications for the next version of the draft.
Roche	General	Section 5.3	Roche appreciates the reference to different methodological approaches to estimate and evaluate meaningful change on COAs including anchor-based, distribution-based and Cumulative Distribution Function plots. Roche proposes adding text to acknowledge that "**the level of change that is meaningful on a PRO or other COA type is dependent on the condition, patient population and associated demographic characteristics of the sample (e.g., see Wang et al. 2011 and Mouelhi et al. 2020).**" References: Wang, Y. C., Hart, D. L., Stratford, P. W., & Mioduski, J. E. (2011). Baseline dependency of minimal clinically important improvement. Physical therapy, 91(5), 675-688.	We agree MIDs can be subject to issues such as baseline dependency. However, there is a balance to find within the guideline between providing general guidance for assessors and not being a statistical textbook. We will consider is such a clarification is needed for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Mouelhi, Y., Jouve, E., Castelli, C., & Gentile, S. (2020). How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. Health and Quality of Life Outcomes, 18(1), 1-17.	
Roche	General	Section 3.1 & 3.2	Given the variation in which different concepts are used (e.g. "final outcomes", "determinant outcomes", "patient-centred outcomes", "standardised set of outcomes" & "core outcomes set"), can EUnetHTA please provide definitions or clarification on how these terms are being used or applied?	We will consider if clarifications are needed for the next version of the draft.
HTAi PCIG		Line 211	This section discusses the definition of patient-centred outcomes without any reference to patients being consulted or involved in this process. PROs may not be important outcomes for some conditions or populations. The document also needs to make clear that sometimes health state utility values (HSUV) alongside PRO measures may be needed to reduce uncertainty. In recent draft FDA guidance on Patient-Focused Drug Development: Selecting, Developing, or Modifying Fit-for Purpose Clinical Outcome Assessments (https://www.fda.gov/media/159500/download), the guidance highlights the need for outcomes important to patients: Section starting at Line 209 of the draft FDA Guidance: *In a clinical trial, it is important to carefully select concepts that, when measured adequately:* • ***Reflect an aspect of health that is important to patients*** • *Have the ability to be modified by the investigational treatment* • *Could demonstrate clinically meaningful differences between study arms within the time frame of the planned clinical trial* To achieve this, see line 40 of the draft FDA guidance: *The "FDA recommends that stakeholders **engage 41 with patients** and other appropriate subject matter experts (e.g., qualitative researchers, clinical 42 and disease experts, survey methodologists, statisticians, psychometricians, patient 43 preference researchers) when designing and implementing studies..."*	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Mihai Rotaru, EFPIA	21-22	Appendix A	**Appendix – Duration of Response** Proposed change: ***"Duration of response (DoR) is defined as the time from randomization to disease progression or death in patients that achieve a complete or partial response. This outcome is of use when assessing interventions that have evidence of durable responses in a meaningful portion of the trial population."*** Rationale: EFPIA recommends the inclusion of duration of response (DoR) to the list of outcomes that are used to capture the effect of cancer therapies. EFPIA believe DoR will prove a useful clinical and patient-relevant outcome in future JCAs of cancer treatments and ATMPs.	We will consider if this addition is necessary for the next version of the draft.
Mihai Rotaru, EFPIA	21-22	Appendix A	**Appendix – Progression-free survival 2** Proposed change: ***"Progression-free survival 2 (PFS2) is defined as the time from randomisation (or registration, in non-randomised trials) to second objective disease progression, or death from any cause, whichever first. Patients alive and for whom a second objective disease progression has not been observed should be censored at the last time known to be alive and without second objective disease progression."*** Rationale: EFPIA recommends the inclusion of PFS2 (randomization to second disease progression or death as defined in EMA "Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man") in addition to OS as an outcome measure in oncology. Due to the challenges with obtaining OS data mentioned in Appendix A, PFS2 is increasingly used to assess long-term outcomes of an intervention. References: 1. EMA/CHMP/27994/2008/Rev.1. Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man.	We will consider if this addition is necessary for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			(https://www.ema.europa.eu/en/documents/scientific-guideline/appendix-1-guideline-evaluation-anticancer-medicinal-products-man-methodological-consideration-using_en.pdf)	
Tanja Podkonjak, Takeda	21-22	Appendix A	Line 203-205 of the current guideline states the following: *"For example, hazard ratios estimated using a Cox model require that the proportional hazards assumption approximately holds. If not, hazard ratios are not valid estimates and another effect measure should be used, such as the restricted mean survival time."* We note that the guideline does not provide a definition of restricted mean survival time. Takeda requests the guideline be updated to include and provide a definition of restricted mean survival time if proportional hazards assumption is not valid, either in Appendix A or earlier in the document.	This sentence that was only used as an example will be removed from the guideline.
François Houÿez, Eurordis	12-13	346-381	This section focuses on adverse events, which can be defined as "Any untoward medical occurrence in a patient or clinical trial subject administered a medicinal product and which does not necessarily have a causal relationship with this treatment" as per Dir 2001/20/EC Art 2. The same definition can apply both to pharmaceuticals and to other technologies. However, it focuses on events that have occurred, and this minimises the issues a technology can raise. Adverse events defined as identified events (that have occurred) only partially describe the dangerousness of a technology. The dangerousness can also be described by potential risks, with no occurrences. For example, a medicine such as thalidomide exposes to very high risks of teratogenicity – even if no case has occurred in the EU since its reintroduction on the market in 2007, thanks to pharmacovigilance activities, including effective risk minimisation measures (Prevention Prevention Programme). For thalidomide, the risk is high and yet no AE are reported in relation to teratogenicity. To really characterise the dangerousness of a technology, both its identified effects (AEs) and its unidentified but potential risks need to be considered. This can have a major impact on the organisational aspects to be assessed at the national level (for thalidomide, organisations aspects include the feasibility and acceptability of the Pregnancy Prevention Programme). The potential character of a risk can be proposed based on animal or in vitro	We think that this kind of risk (already known to be associated with the drug but not observed in the trial, like thalidomide for example) are already described in the SmPC. There is no need to duplicate all the safety information from the SmPC to the JCA. These documents are complementary.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			or in silicio studies.	
Marjorie Morrison, Lymphoma Coalition	Primarily 9-10 General as noted.	252-285	**COS: Core Outcome Set** The concept of standardised and improved outcome reporting across a disease using a core outcome set, or set of outcomes, has been studied in relation to cancers and oncology trials. Studies point to the integration of COS as an integral component to optimising clinical trial design, further facilitating market access and ensuring trials where evidence is challenging in smaller patients sizes. This supports the objective of improving patient outcomes while harmonising and reducing differents in measured outcomes. Variances between outcome measures and clinical trials, as well as those between defined hematological malignancy entities, may occur. Therefore, a commonly accepted and standardised set of outcomes measured and reported in trials should clearly address the needs of all stakeholders to collaboratively define COS. The COS provides the "minimum outcomes set that should be collected in further clinical trials on a given condition" with COS improving both the "comparability of clinical trials and consistency in reporting, reduce selective reportingbias and ensure that appropriate outcomes valued by a range of stakeholders are measured." (2) **COS: HARMONY Consortium (General)** The Healthcare Alliance for Resourceful Medicine Offensive against Neoplasms in Hematology (HARMONY) was established in 2017 and currently engages approximately 90 organisations/partners across 22 different European countries. HARMONY proposed a platform to define COS for HMs using the Delphi survey method where the study and final consensus meeting concluded that COS could be defined and applicable to all hematological maligancies and used to refine COS and apply witin future clinical trials to reduce bias in outcome reporting, contribuitng to improved patient management and clinical patient care. (1)	Thank you for your comment. This section has been modified.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			As further studies of traditional clinical vidence data sources and outcomes by domain also suggest there is a need for caution with respect to the gaps in evidence prposing that "fewer evidence sources and earlier phase studies used to meet HTA requirements" may, in fact, lead to negative decisions, there are key factors for consideration with respect to COS and the EUnetHTA21 deliverable addressing endpoints. (1) Additionally - with respect to future data considerations, the outcomes of COS (or rather the definition and application of COS) will no doubt play a role in determining what future data are collected and analysed. The process and results of defining COS has signficiant implications to future data sets in terms of what information is available. This is of particular importance when considering data collection and reporting for rare lymphomas. **COS: Real-World Evidence. (Page 9/267-272)** In addition to application to clinical setting and/or incorporation into clinical guidelines or clinical practices, COS should be considered as essential to observational studies to support real-world evidence research. More specifically, consideration of real-world evidence alongside clinical evidence should be based on the inclusion of real-world data and patient-reported outcomes (PROs) in clinical trial design, especially given the implications of COS beyond trials to clinical patient care in a real-world environment. **COS: Novel Outcome Measures. (General)** Unquestionably, quality-of-life outcomes are of significance importance to patients. Conceivably, a percentage of patients may be more interested in managing quality-of-life issues that impact their respective day-to-day function rather than a core focus on traditional outcomes, namely improvement in relation to overall survival.	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			The inclusion of novel outcome measures will likely provide relevant patient information more rapidly than traditional outcomes and as a result, may increase availability for patients who require novel treatments or interventions. (3) **We are in full agreement that to support the definition of COS, the inclusion of patient organisations as a key stakeholder is essential. In addition, the importance of patient involvement as a key stakeholder will further support the inclusion of patient-reported outcomes for the purpose of better understanding efficacy and ensuring that endpoints are those that are most meaningful to patients; support data harmonisation, and enhance measurement in trials and observational studies.**	
Tanja Podkonjak, Takeda	21-22	683-684, Appendix A general	Current text: *"In oncology most often reported disease related outcomes are progression free survival (PFS) as surrogate for OS, event free survival (EFS), or disease-free survival (DFS)."* Proposed text: *"In oncology most often reported disease related outcomes are progression free survival (PFS) ~~as surrogate for OS,~~* ***PFS2 as a surrogate for OS****, event free survival (EFS), or disease-free survival (DFS)."* Rationale: PFS is a clinically important outcome as a standalone endpoint and is not only included in oncology studies as a surrogate for OS. This position is supported by the EUnetHTA 2015 guideline on *Endpoints used for Relative Effectiveness Assessment: Clinical Endpoints*[1], a guideline referenced in the proposed document, which states the following on page 10: "***In oncology, PFS is an intermediate endpoint that is relevant on its own right."*** A PFS event marks a progression or loss of response to an intervention which has prognostic implications (i.e., if a patient is refractory due to	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			treatment) as well as important clinical consequences such as informing the termination of existing treatment and the initiation and choice of any subsequent therapy. The current text is misleading as it implies that PFS is only reported as surrogate for OS, which is unfounded. Furthermore, we request the guideline include PFS2 (randomization to second PD or death as defined in EMA "Appendix 1 to the guideline on the evaluation of anticancer medicinal products") as an outcome measure in Oncology. Due to the challenges with obtaining OS data mentioned in Appendix A, in particular for conditions with long durations of survival, PFS2 is increasingly used to assess long-term outcomes of an intervention. Takeda recommends the guideline be updated to include PFS2 in the Oncology Appendix as a common reported disease related outcome in oncology and provides a supporting paragraph on PFS2 as an endpoint. Reference: 5. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Clinical Endpoints. Adapted version 2015	
Bayer	9-10	3.2	Proposal: Suggest omitting section as there is no standard for all indications and the core data set has not been developed from HTA perspective. This section might be misleading.	Thank you. This section has been modified.
S.Walleser Autiero Medtronic	10-11	3.3. Surrogate endpoints	Evidence from previous studies showing well established links between surrogates and LT patient outcomes should also be accepted as part of the submission by HTDs. Consideration of a Linked Data Approach for these associations should be included. Extension of acceptability of surrogate endpoints to endpoints accepted by the international scientific society should be considered, as these have the best understanding of the latest evidence, and most relevant clinical expertise of the specific disease and therapy area.	Thank you. The level of acceptance of surrogate endppoints for HTA purposes differs between MS.
M. Ermisch – GKV-	12-13	Sec. 4.3	The reporting on adverse events need to be refined. In addition to the report on the grading scale for severity of adverse events used in the study, overall rates of severe adverse events (i.e. e.g. CTCAE ≥ 3) need to be reported.	Overall rates of SAE is already included (see box line 380). Reporting using both SOC and PT will lead to unreadable JCA report, due to this very

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
SV			In addition to the report of overall event rates, analyses that are more detailed need to be included: adverse events, serious adverse events and severe adverse events should be reported differentiated according to system organ classes (SOC) and preferred terms (PT). This report can be limited to events that either occurred in more than 10% of patients in at least one treatment arm, or (for severe and serious events) that occurred in more than 5% of patients in at least one treatment arm, or that occurred in at least 10 patients equalling at least 1% of patients in one treatment arm. It is not sufficient that specific adverse events (PT / SOC) are only presented at the request of a member state. The ´profile´ of adverse events can rarely be predicted in advance. It would be a completely selective request. Therefore, in the box "if applicable" should be deleted.	specific MedDRA classification. No change.
Roche	7-8	209 (Box)	Current wording: " - Effect measures should not be specified by MS. The HTD is responsible for presenting results using appropriate effect measures in accordance with good clinical and statistical practice. - If an MS still wants to specify an effect measure, the wording should follow this template: "[Outcome of interest] with treatment effect expressed preferably as [insert effect measure]". " Proposed wording: " - Effect measures should not be specified by MS. The HTD is responsible for presenting results using appropriate effect measures in accordance with good clinical and statistical practice. ~~**- If an MS still wants to specify an effect measure, the wording should follow this template: "[Outcome of interest] with treatment effect expressed preferably as [insert effect measure]".**~~ "	We understand the point but according to the HTAR MS can request outcomes they see fit and therefore this possibility must be taken into account.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Rationale: The Guideline should aim at internal consistency and, therefore, avoid conflicting recommendations. We suggest dropping the second bullet here, which is not consistent with the previous one.	
Matias Olsen, EUCOPE	5 - 6	118–123	Articles that are relevant for the use of surrogate endpoints should be included as reference in the guidance. Recital 24 of the Regulation (EU) 2021/2282 on health technology assessment specifies that methodologies for 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, e.g. OMPs and ATMPs while Article 4 (1) specifies that the Coordination Group, which shall ensure that the joint work carried out is of the highest quality, shall consider the specificities of the health technology, including OMPs and ATMPs when developing the procedures for joint work. Add: “ Articles from Regulation (EU) 2021/2282 directly relevant to the content of this particular guideline are: • Recital 2, • **Recital 24** • Recital 28, • **Article 4 (1): Consideration of specificities of certain technologies, including orphan medicinal products, vaccines and advanced therapy medicinal products,** • Article 8: Initiation of joint clinical assessments, • Article 9: Joint clinical assessment reports and the dossier of the health technology developer, • Article 13: Member States’ rights and obligations “	We will consider if this is necessary for the next version of the draft.
Mihai Rotaru, EFPIA	4	73	List of Acronyms: Additional abbreviations will be needed on final proposal publication e.g. COA, Clinical Outcomes Assessment.	It will be revised in the next version of the draft.

EUnetHTA 21 Public Consultation Comments and Responses
Of D4.4 - Endpoints

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Hayley Chapman, PFMD	4	73	Include acronym for CA – Clinical Assessment	COA will be added.
Edwards Lifesciences	5	114-115/ Section 1.1 Problem statement, scope and objectives	In our perspective effectiveness and efficacy are interlinked but they are not interchangeable, and this is even more true in the context of a health technology assessment. Per the HTA regulation "*Health technology assessment (HTA) is a scientific evidence-based process that allows competent authorities to determine the* ***relative effectiveness*** *of new or existing health technologies*". Therefore using "effectiveness to describe efficacy or effectiveness" creates a lot of confusion and weaves and contradicts the specifics of the JCA requirements (Article 2(6) of the HTA Regulation on the Definition of JCA) and any other regulatory evidence requirements.	Efficacy and effectiveness are not equivalent terms but for readability it was simpler to stick to one only and we do not think this decision has a major impact on the content of the guideline.
Edwards Lifesciences	5	115-116/ Section 1.1 Problem statement, scope and objectives	By using very generic terms like "treatment", "intervention" and "health technology" for any health technology that can be assessed, weaves the specifics and the different needs for MD and for pharmaceutical products. The HTA regulation clearly differentiates between the specifics of medical devices, and pharmaceutical treatments, hence we believe that the methodologies should also capture the specifics of the categories of treatments in the same line.	Clinical research methodology is more standardized and harmonized in the realm of pharmaceutical products than MD. We agree this is reflected within the document. Nonetheless, we believe this general guidance will help assessors for different types of health technologies.
Matias Olsen, EUCOPE	5	86-98	While interpretation of the clinical relevance of effect measures (and thus the added value of a new technology) may differ based on the current standard of care, the interpretation of relevance of surrogate endpoints is based on scientific grounds, i.e. correlation of effects and thus their relevance should in principle not differ between member states.	Correlation / association are not on/off concepts. They come in degrees and are usually assessed by multiple continuous indices. It is therefore understandable the appraisal of what constitutes sufficient validity of a surrogate endpoint can vary.
Matias Olsen, EUCOPE	5	80-85	As we have noted in our feedback on *inter alia* draft guidelines D4.2 "Scoping Guideline" and D7.1.1. "practical guideline on the interaction between HTD and HTAb", it would, in the interest of developing a robust submission dossier that can support a successful EU HTA procedure, be appropriate if the scoping process allows for an open discussion and alignment between the assessors and developers on appropriate elements of the PICO. Especially for rare	This is out of scope of the guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			diseases, and large heterogeneity regarding standards of care between Member States, there should be a discussion meeting with the HTD prior to the definition of the final PICO(s), i.e. a scoping meeting. Special mention should be made to outcomes used for regulatory approval. These should always be included in the list of outcomes to be assessed as part of the PICO scheme.	
Matias Olsen, EUCOPE	5	100-105	The objective of this guideline should include the aim to ensure consistency in the outcomes that are requested from developers by the Member States and that a homogeneous assessment of those outcomes is made by all Member States. It is not uncommon under the current fragmented approach to HTA that different Member States do not accept the same outcomes i.e., QoL, surrogate endpoints, etc. To ensure a high-quality joint clinical assessment there needs to be a homogeneous approach to the choice of outcomes across all EU. Its analysis should also be homogeneous, it is not conducive to the use of future joint clinical assessments that a Member State places importance to QoL while another Member State could completely disregard it. Replace: "The objectives of this guideline are twofold. The first objective is to provide guidance for MS in defining relevant outcomes during the scoping process. The second is to help assessors and co-assessors in assessing and reporting all the necessary elements that MS need to carry out for national appraisals of the clinical added value of the health technology." With: "The objectives of this guideline are ~~twofold~~ **threefold**. The first objective is to provide guidance for MS in defining relevant outcomes during the scoping process. The second is to help assessors and co-assessors in assessing and reporting all the necessary elements that MS need to carry out for national appraisals of the clinical added value of the health technology**. The third is to ensure a homogeneous list of outcomes is requested from the HTD by the MS and that a homogeneous assessment of those outcomes is made**	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			by all MS."	
Sebastian Wernervfa	5	85-90	"*While MS are required to give due consideration to the JCA reports published (Article 13 (1)), the clinical relevance or interpretation of the measure of relative effectiveness may differ between MS when drawing conclusions regarding the clinical added value of a treatment at a national level. Therefore, appropriate reporting of the methodological and statistical elements and results of the analyses of the outcomes requested is essential (Article 9(1)).*" The decision what constitutes a clinically relevant outcome or an adequate interpretation of the measures using responder definitions (incl. MID) should be based on scientific rationale that is harmonized across Member States. Member States should use the same responder definitions (incl. MID) that are valid or established according to the generally accepted state of scientific knowledge and the international standards of evidence-based medicine. Member states should only use different responder definitions, if available scientific evidence shows differences in national patients that would lead to different interpretations of the data for the Member State. Thus, the assessment of responder definitions should be based on scientific rationale with well-defined methodological criteria to ensure joint decisions in the JCA report on the validity of responder definitions (incl. MID). The guidance should provide more details on the criteria of acceptability of responder definitions (incl. MID) considering the generally accepted state of scientific knowledge and the international standards of evidence-based medicine. EUnetHTA21 and the CG should use its best endeavours to reach a consensus on acceptability criteria and work towards a harmonized approach.	The guideline aims for a factual reporting of elements pertaining to validity, reliability and interpretability of scales as these concepts are not on/off concepts. They come in degrees. Elements pertaining to appraisals are left at the discretion of the MS.
S.Walleser Autiero Medtronic	5	100-105	One of the objectives of this document is to provide guidance for member states on how to define relevant outcomes, including how they should be measured and at what time points. It should be considered that if MS have different views on the outcomes and their operationalisation, potentially the scope and subsequently submission needs might increase to an extent that would make it impossible for HTDs to submit all the relevant analyses responding to MS specific needs for outcome measurement in the limited time available. For timely JCA, it should be considered that the final scope provided to the HTD provide a consolidated version of the PICO including	Thank you for this comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			consolidation on what is required in terms of outcome measures.	
Roche	5	101-105	Given the two objectives stated in this section, and that "all the requirements for reporting and assessment mentioned in this guideline suggest that HTDs are supposed to present the necessary elements in their submission dossiers" ; when taken together, imply that there will not be JCA dossier submission guidance for HTDs. Roche would like to seek clarity on whether guidance will be provided for HTDs, whilst also noting the importance of including HTDs in the JCA process.	There will be a guideline and template for the submission dossier.
Mihai Rotaru, EFPIA	5	100-103	Current wording: *"The objectives of this guideline are twofold. The first objective is to provide guidance for MS in defining relevant outcomes during the scoping process. The second is to help assessors and co-assessors in assessing and reporting all the necessary elements that MS need to carry out for national appraisal of the clinical added value of a health technology."* EFPIA suggests addition of the following sentence: ***"This guideline should be interpreted taking the Member State context into account for appraisal, as well as other important dimensions, for example, the disease context (including public health considerations) or type of intervention including input from patient and clinical experts on outcomes relevant for the condition and place in therapy***." Rationale: In addition to MS autonomy in selecting outcomes for their national appraisal, there are also other dimensions to consider, for example the disease context (rare disease, progression type, diagnosis…) when selecting appropriate endpoints. EFPIA suggest clinical and patient relevance and disease context, including the public health perspective, be considered via consultation with patient and clinical experts,	This guideline is not intended to define national procedure, which should remain at MS discretion only.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			and alignment with outcomes considered in relevant clinical guidelines. Early engagement with the HTD can help provide these important contexts, to the benefit of the overall process.	
EFSPI	5	85-88	The judgement what constitutes a clinically relevant outcome, responder definition, etc. should be harmonized across member states (MS) based on a scientific rationale. It is difficult to understand if one member state considers effect size estimates, say Cohen´s d of 0.2 as clinically relevant and another MS 0.25 and a third MS 0.28 as clinically relevant, or if MS consider different MIDs for PROs are relevant for decision making. **Current wording**: "While MS are required to 85 give due consideration to the JCA reports published (Article 13 (1)), the clinical relevance or 86 interpretation of the measure of relative effectiveness may differ between MS when drawing conclusions 87 regarding the clinical added value of a treatment at a national level." **Proposed wording**: "While MS are required to give due consideration to the JCA reports published (Article 13 (1)), the rating of the additional benefit of a treatment may differ at a national level which is based on the clinical relevance of the measure of relative effectiveness."	Already addressed issue.
	5	96-99	The guideline says that "the conclusions that MS can draw regarding the clinical added value of a treatment can be impacted by factors such as appraisal of the validity and reliability of the measurement scales of instruments or of the relevance of intermediate or surrogate outcomes." The assessment of the validity and reliability of the measurement scales of instruments should be based on a scientific basis and should be harmonized across member states. Please ensure that such a common appraisal will not to be performed by individual member states. **We propose to remove this sentence.**	Based on scientific rationale or evidence-based do not imply that decision-making is devoid of any kind of appraisal or debate. Validity, reliability and interpretability of either outcome measurement instruments or surrogate outcomes are not on/off concepts. They come in degrees. This is why the guideline aims for a factual reporting of the methods and results elements regarding these aspects, but MS can appraise them the way they see fit.
Sebastian Werner vfa	5	96-99	*"However, the conclusions that MS can draw regarding the clinical added value of a treatment can be impacted by factors such as appraisal of the validity and reliability of the measurement scales of instruments or of the relevance of intermediate or surrogate outcomes."*	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			The assessment of the validity and reliability of the measurement scales of instruments are of scientific nature and should be based on well-defined methodological criteria to ensure joint decisions in the JCA report. Therefore, the guidance should provide more details on the criteria of acceptability of measurement scales considering the generally accepted state of scientific knowledge and the international standards of evidence-based medicine. EUnetHTA21 and the CG should use its best endeavours to reach a consensus on acceptability criteria and work towards a harmonized approach.	
Mihai Rotaru, EFPIA	5	77-79	Proposed change: *Clinical outcome assessment**s are used in clinical trials and are*** ~~*is*~~ *a key component of health technology assessment (HTA).* ~~***It is***~~ ***They provide measures*** *of the clinical benefit of the targeted treatment on patient-centred outcomes* ***and include patient reported outcomes, clinician reported outcomes, patient generated health data outcomes, observer reported outcomes and performance outcomes[1]*** *(see the definition in Section 3.1* ***for more information****)* Rationale: Clinical outcome assessment is a broad term that includes the four categories of outcome measures that have been referenced in the proposed wording. EFPIA feel it is important to make this clear in the introduction to ensure MS are aware of the breadth of clinical outcome assessment when requesting outcomes of interest during the PICO process. Reference: Walton MK, Powers JH, Hobart J et al. ISPOR Task Force Report Clinical Outcomes Assessment: Conceptual Foundation – Report of the ISPOR Clinical Outcomes Assessment – Emerging Good Practice for Outcomes Research Task Force. Value in Health 2015; 18; 741-752 - DOI: 10.1016/j.jval.2015.08.006	This modification will be considered for the next version of the draft.
S.Walleser Autiero	5	104-105	HTD are supposed to present the necessary elements in their submission dossiers as far as they are able to do so. Depending on the timing of the HTA it can be difficult to do that as some data on relevant outcomes might not be	Indeed. But, this issue is out of the scope of this guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Medtronic			available at the time of submission in particular if the timing of the assessment is chosen based on CE mark but not on availability of HTA relevant clinical evidence (or not publicly available and bound to academic confidence).	
Tanja Podkonjak, Takeda	5	106-107	Current text: *"In the context of JCA, outcomes cannot be dissociated from the way in which they are statistically analysed."* Proposed text: *"In the context of JCA, outcomes cannot be dissociated from the way in which they are statistically analysed as well as their patient and clinical relevance."* Rationale: Statistical methods and rigour are important in interperting the results of a treatment, however, the relevance of outcomes to patients as well as clinical decision making is also critical in understanding the relevant measure of efficacy for a treatment. Takeda suggests the text be modified to include the relevance of an outcome to patients and clinical practice.	This sentence aims at introducing the other guidelines which deals with methodological and statistical aspects. It does not mean these are the only aspects to be considered.
Roche	5	86	This document should provide additional guidance on the harmonization of evidence required at the pan European level and reflect a more centralized perspective. The current guideline provides many statements (such as "the clinical relevance or interpretation of the measure of relative effectiveness may differ between MS when drawing conclusions regarding the clinical added value of a treatment at a national level") that reference the decentralized decision making at the member state level. We suggest that the guidelines apply a broader, more European level scientific understanding of the evidence including flexibility on the methods. The deliverable D4.4 should provide clear scientific guidance to assessors/co-assessors and HTDs applicable at the European level, while ensuring sufficient flexibility to account for therapeutic area and	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			technology specific considerations.	
HTAi PCIG	5	Line 114	The document states that "*effectiveness is the term used to describe efficacy or effectiveness*". As there is a clear difference in the type of evidence used to prove efficacy or effectiveness it is an imprecise approach to use the word effectiveness for both and may lead to confusions in the adoption by different member states. This risk needs to be addressed.	Already addressed issue.
MTE	5	118	Relevant for medical devices to add recital 38. HTA-R work should be separate and distinct from the regulatory assessments conducted pursuant to Regulations (EU) 2017/745 and (EU) 2017/746.	It will be considered for the next version of the draft.
Hayley Chapman, PFMD	6	147-177	This document refers to COAs. It would be helpful to the patient community and industry to communicate what consideration has been given to other methods including real world data capturing patients' or patient preference, and why those methods may or may not be appropriate for consideration. Several regulatory agencies have recently released guidelines on this (see examples below) and the HTA process should also provide similar recommendations in order to harmonize and streamline efforts between regulatory and HTA agencies. https://www.fda.gov/about-fda/cdrh-patient-science-and-engagement-program/patient-preference-information-ppi-medical-device-decision-making https://www.fda.gov/regulatory-information/search-fda-guidance-documents/submitting-documents-using-real-world-data-and-real-world-evidence-fda-drug-and-biological-products https://www.gov.uk/government/publications/mhra-guidance-on-the-use-of-real-world-data-in-clinical-studies-to-support-regulatory-decisions	Real-world data use is dealt within the D4.6 validity of clinical studies guideline. Digital acquisition of patient generated data in real world settings use will be briefly expanded in the next version of the guideline.
Mihai Rotaru, EFPIA	6	147-174	The categorisation of outcomes in these three main categories may miss important considerations; for example, where does mortality fit within this grouping? Patient experience or patient acceptability of new technology are missing. Connected	Mortality is a clinically reported outcome as death must be usually asserted by a physician in many countries.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			digital health technologies can offer patient experience data and are not only connected to compliance or improved adherence to medication. EFPIA also encourages EUnetHTA to align the Endpoint guidance and classification with other international guidelines including regulatory from EMA and FDA (e.g. https://www.fda.gov/about-fda/clinical-outcome-assessment-coa-frequently-asked-questions, https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines/clinical-efficacy-safety-guidelines) medical societies and international guidelines. This is in line with the scope of the objectives expressed in the D4.4 Endpoint project plan.	Experience and acceptability are not dealt within the guideline as we do not consider these are primary outcomes of interests in the context of HTA. We agree the use of digital health technologies must be better addressed in the next version of the draft.
Natacha Bolanos, Lymphoma Coalition	6	152-165	The definition of PROs should be more well elaborated to include: *"PRO data should be considered along with cost and survival data when approving new therapies. PRO data can also be helpful when assessing existing treatment options for patients, particularly for drugs with minor outcome and toxicity differences. PROs capture patient reports of outcomes in individual domains without providing information about patient preferences across domains. The relative importance of these domains is quantified by patient preference methods. PRO measures are aligned with societal values, and broadening the definition of society to extend beyond national boundaries"* Likewise, it could be complemented with additions referred to how PRO are valuable for health care and health economic decision-making at institutional, regional, national, and international levels. PROs can be useful in reimbursement algorithms to ensure delivery of quality cancer care in value-based financing environments.	We do not think such unspecific statements (and sometimes vague such as "*are aligned with societal values, and broadening the definition of society to extend beyond national boundaries"*) will improve the definition of what is a PRO in the context of this draft. We will briefly mention PROMs can be used to estimate utility values, but economic evaluation of health technologies is out of the scope of JCAs which are primarily concerned by relative effectiveness and safety.
Mihai Rotaru, EFPIA	6	166-174	Proposed change: *"Performance outcomes are ~~close to~~ clinically reported outcomes ~~but~~ **that** require active patient involvement **to complete a standardised task (e.g., 25-foot walk test with an ankle-worn sensor, cognitive tests) [FDA-NIH Biomarker Working Group. BEST]**~~, (e.g., tests of walking, cognitive tests)~~. There is also increasing use of **digital***	Thank you. These propositions will be considered for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			health technologies to passively monitor normal daily activities or behaviour (e.g., use of a wrist-worn sensor to measure daily step count) [Huhn S, Axt M, Gunga HC, et al. 2022], ~~*patient-generated health data such as outcomes using connected digital health technologies (e.g., monitoring devices for medical adherence)*~~*. These devices can allow an automated measure of outcomes in settings other than the usual visits for clinical studies, such as in home settings [2,10].* ***Digital health technologies may also be used to collect other types of patient-generated health data such as medication adherence****. Use of such technologies could lead to …."* Rationale: EFPIA proposes editing the text relating to Performance Outcomes to clarify that these can be measured actively via traditional or digital health technologies as well as highlighting an emerging category of COAs relating to passive monitoring.	
Matias Olsen, EUCOPE	6	147-152	It would be helpful to include the concept of anatomic endpoints here as an example of a clinician-reported outcome, e.g. an anatomic feature that is measured at the end of a study to assess whether a given treatment works.	Describing in depths what are all the subtypes of clinically reported outcomes is not the main purpose of the guideline.
EFSPI	6	161-165	EFSPI considers observer-reported outcomes (ObsROs) to constitute another main, and separate, source of information. This is of particular relevance in the context of paediatric studies where the patient may be too young to self report and in the context of neurodevelopmental or neurodegenerative disease areas where the patient may be cognitively impaired, making a self report unreliable. We propose a new paragraph to start after the definition of PRO is complete (line 161) and the definition provided in the EMA Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man can be used. We additionally propose that a distinction should be made between an ObsRO and a proxy-reported measure by referring to the EMA Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. **Current wording**: "can be assessed using PROMs." **Proposed wording**: "can be assessed using PROMs.	The inclusion of ObsROs has already been considered. The proposed rationale seem sound. We will consider it for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Another source of information can be observers. An observer-reported outcome (ObsRO) is a measurement based on an observation by someone other than the patient or a health professional. This may be a parent, spouse, or other non-clinical caregiver who is in a position to regularly observe and report on a specific aspect of the patient's health. An ObsRO measure does not include medical judgment or interpretation. Generally, ObsROs are reported by a parent, caregiver, or someone who observes the patient in daily life. For patients who cannot respond for themselves (e.g., infants or cognitively impaired), we encourage observer reports that include only those events or behaviours that can be observed. As an example, observers cannot validly report an infant's pain intensity (a symptom) but can report infant behaviour thought to be caused by pain (e.g., crying). A distinction should be made between an ObsRO and a proxy-reported outcome. A proxy-reported outcome is a measurement based on a report by someone other than the patient reporting as if he or she is the patient. A proxy-reported outcome is not a PRO. A proxy report also is different from an observer report where the observer (e.g., clinician or caregiver), in addition to reporting his or her observation, may interpret or give an opinion based on the observation." Reference: European Medicines Agency. Appendix 2 to the guidance on the evaluation of anticancer medicinal products in man, 1 April 2016. Accessed 6th October 2022: https://www.ema.europa.eu/en/documents/other/appendix-2-guideline-evaluation-anticancer-medicinal-products-man_en.pdf	
Mihai Rotaru, EFPIA	6	140-143	Proposed change: *"Thus, effect measures are primarily understood as a comparison of the measure of outcomes between* ***one or more*** ~~*two*~~ *interventions groups* ***and/or a control group (placebo or no intervention).*** *Broadly, effect measures are either difference measures (e.g., mean difference in change, risk difference) or ratio measures (e.g., risk ratio, odds ratio, hazard ratio). However, other statistics can be used to express other aspects of a treatment effect such as* ~~*the absolute effect or a*~~ *within-group change [5]."* Rationale: Comparisons might not only be made between two interventions groups, but also	We will consider if a clarification is needed for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			between an intervention and a "no intervention" group (placebo). For an example, in the context of vaccines, outcomes can also be measured "between an intervention group (i.e., vaccination) and a control group (no vaccination or vaccination with a placebo or control vaccine). Also, as currently written, "absolute effect" may be interpreted as one type of difference measures, hence EFPIA recommends rewording.	
Mihai Rotaru, EFPIA	6	149-152	Proposed change: *"Clinically reported outcomes **(ClinRO)** are assessed by healthcare professionals during clinical examination of a patient and involve clinical judgments of patients' observable signs~~,~~ behaviours or other physical manifestations. Technologically assessed outcomes **(also referred to as biomarker data) may be used together with ClinROs to report particular clinical findings or clinical events** ~~require the use of technology such as laboratory tests or medical imaging~~."* Rationale: The distinction between two types of clinician-reported outcomes is not a COA standard. The definition for "Clinically reported outcomes" is the standard definition for clinician-reported outcomes. Technology assessed outcomes is an ambiguous term as it may also include called biomarker data and digitally assessed Performance outcomes such as step counts. The reference used states the following: "Clinically reported outcomes (ClinROs) are evaluations from a trained professional after observation of a patient's health condition and involve clinical judgment or interpretation of the observable signs, behaviours, or other physical manifestations. Performance outcomes (PerfOs), which require patient cooperation and motivation, include tests of walking, dexterity, and cognition. Key outcomes that are assessed using technology include pulmonary and cardiac function tests and neuroimaging for neurological conditions."	Clarifications on the main types of outcomes will be considered for the next version of the draft.
Mihai Rotaru,	6	158-161	Proposed change: *"Some ~~PROMS~~ **PROMs** measure health status (for instance, the EQ-5D instrument*	It will be corrected in the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
EFPIA			*measures health status as a combination of five broad concepts). Other outcomes such as symptoms **(including fatigue and pain)**, anxiety, depression, functioning, impairment, disability and impact on daily living can **also** be assessed using PROMs."* Rationale: Grammatical suggestion - change "PROMS" into "PROMs" and suggest adding "also" in the sentence. Fatigue and pain are symptoms.	
EFSPI	6	166-169	EFSPI proposes editing the text relating to Performance Outcomes to clarify that these can be measured actively via traditional or digital health technologies as well as highlighting an emerging category of outcomes relating to passive monitoring **Current wording**: "Performance outcomes are close to clinically reported outcomes but require active patient involvement, (e.g., tests of walking, cognitive tests). There is also increasing 167 use of patient-generated health data such as outcomes using connected digital health technologies 168 (e.g., monitoring devices for medical adherence). These devices can allow an automated measure of outcomes in settings other than the usual visits for clinical studies, such as in home settings [2,10]. **Proposed wording**: "Performance outcomes are close to clinically reported outcomes but require active patient involvement, to complete a standardised task (e.g., 25-foot walk test with an ankle-worn sensor, cognitive tests) [FDA-NIH Biomarker Working Group. BEST]. There is also increasing use of digital health technologies to passively monitor normal daily activities or behaviour (e.g., use of a wrist-worn sensor to measure daily step count) [Huhn S, Axt M, Gunga HC, et al. 2022]. These technologies can allow an automated measure of outcomes in settings other than the usual visits for clinical studies, such as in home settings. Digital health technologies may also be used to collect other types of patient-generated health data such as medication adherence."	Clarifications about PerfOs and digital health outcomes will be considered for the next evrsion of the draft.
Roche	6	166-169	Roche proposes editing the text relating to Performance Outcomes (PerfO) to clarify that these can be measured actively via traditional or digital health	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			technologies as well as highlighting an emerging category of COAs relating to passive monitoring: "Performance outcomes are close to clinically reported outcomes **but require active patient involvement, to complete a standardised task (e.g., 25-foot walk test with an ankle-worn sensor, cognitive tests) [FDA-NIH Biomarker Working Group. BEST]. There is also increasing use of digital health technologies to passively monitor normal daily activities or behaviour (e.g., use of a wrist-worn sensor to measure daily step count) [Huhn S, Axt M, Gunga HC, et al. 2022].** These technologies can allow an automated measure of outcomes in settings other than the usual visits for clinical studies, such as in home settings. **Digital health technologies may also be used to collect other types of patient-generated health data such as medication adherence**. Use of such technologies could lead….." References: *FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring, MD: US Food and Drug Administration–US National Institutes of Health; 2021.* *Huhn S, Axt M, Gunga HC, et al. The impact of wearable technologies in health research: scoping review. JMIR mHealth uHealth 2022;10(1):e34384.*	
Mihai Rotaru, EFPIA	6	130-132	Proposed change: *"If the outcome is pain, the measure of the outcome could be "change in the level of pain on a patient-reported **numeric rating scale (from 0-10)** ~~visual analogue scale of 100 mm~~ at 24 hours after initiation of the treatment".* Rationale: Update reference to current standards. The visual analogue scale for Pain has become less utilised in pain research, instead, recommend using the numeric rating scale.	It will be corrected. Thank you.

EUnetHTA 21 Public Consultation Comments and Responses
Of D4.4 - Endpoints

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Mihai Rotaru, EFPIA	6	163-165	Proposed change: *"These cases are referred to as ~~PROs~~* ***Observer Reported Outcomes*** *~~answered by "proxies~~". This distinction is important because the person who is assessing the outcome can impact the accuracy of the information.* ***Proxy report is to be distinguished from observer-reported outcome (ObsRO), which is a measurement based on a report of observable signs, events or behaviours related to a patients' health condition by someone other than the patient or health professional. Generally, ObsROs are reported by a parent, caregiver, or someone who observes the patient in daily life****."* Rationale: As stated above, PROs and ObsRO should be distinguished. Only Patient-Reported Outcomes are PROs. Information collected from Observers is categorised as ObsRO and should be limited to reporting on what the observer can actually observe. Observers are not supposed to report on behalf of the patients. Reference: BEST glossary https://www.ncbi.nlm.nih.gov/books/NBK338448/#IX-C	Clarifications will be made in the next version of the draft.
Mihai Rotaru, EFPIA	6	126-127	Proposed change: *"Outcome" is any concept that can be used for estimating treatment effectiveness, such as mortality, remission,* ***disease control****, health-related quality of life (HRQoL),* ***function, prevention of an event,*** *symptoms,* ***observer-reported outcomes****,* ***clinical-reported outcomes and performance outcomes****, as well as safety."* Rationale: The definition of outcome is limited, and needs to be broader, recognising clinical endpoints (disease control) and observer-reported outcomes. Function and prevention of a clinically relevant event (i.e. prevention of infection or disease in the context of vaccines) are very important outcomes for patients, so should be explicitly mentioned in this list. Safety is a different concept to treatment effectiveness.	It will be considered for the next version of the draft.
Matias	6	163-164	We suggest adding a definition of observer-reported outcomes (ObsRO) in	Clarification will be made for the next

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Olsen, EUCOPE			addition to the "proxy" definition for those cases where an observer completes the PRO. Unlike a PRO, the ObsRO is *a priori* planned to be completed by an observer (usually a parent or the caregiver of the patient). Add: "These cases are referred to as PROs answered by "proxies". **Additionally, to answer by proxy instruments can also be established, e.g. in case of very young children, where a parent or the caregiver is *a priori* defined as the person to complete the instrument in question. The measurements resulting from such instruments are referred to as observer-reported outcomes (ObsRO).**".	version of the draft.
Natacha Bolanos, Lymphoma Coalition	6	126-127	The phrase *"treatment effectiveness, such as mortality, remission, health-related quality of life (HRQoL), symptoms and safety"* could be finetuned, to reflect better the measurable elements: functional goals of therapy, the incidence or prevalence of adverse events, measurable residual disease status (or response to treatment rather than remission as this is limited concept), quantifiable impact on quality of life and wellbeing, etc.	At this stage, outcomes are introduced as broad concepts. Operationalizations of outcomes are defined after.
HTAi PCIG	6	Line 170/171	The sentence: *"Use of such technologies could lead to benefits such as better compliance or expansion of participation"* describes an intervention and not an outcome. If it is mentioned here, it should also be discussed how to deal with the fact that a tool used to measure outcomes may be acting as an intervention at the same time.	This is sound. It will be clarified in the next version of the draft.
M. Ermisch – GKV-SV	6	Sec. 2.1	A particular category of outcomes (which might be called "health services (utilization) outcomes" is missing in the definition section. In particular hospitalizations are sometimes used (e.g. in heart failure trials) to assess effectiveness of therapies. These are imperfect outcomes, basically for two reasons: They generally are surrogate measures for morbidity(-events) but are susceptible to provider choices and coding practice and therefore e.g. to performance bias. They will also be influenced by health system characteristics (of MS) and transferability of findings might (depending on the clinical context) thus be limited. The former limitation might be ameliorated by including (e g. in the case of hospitalization) all events (hospitalizations) independent of coded reasons. These limitations should be duly noted and	We have not considered health services utilization outcomes as a patient-centred outcome that is paramount from a HTA perspective.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			discussed.	
GSK	6	Section 2.1	• Please align the definition of 'outcome' with definitions in ICH E9(R1) and regulatory definitions. For example, under ICH E9(R1) a variable/endpoint is measured at the patient level and the variable/endpoint is one of 5 attributes of the estimand that together define the treatment effect of interest. • The text 'the measure of the outcome could be 'proportion of deaths 28 days after inclusion'' is misaligned to ICH E9(R1). The outcome is whether a patient is alive or dead 28 days after inclusion so the measure of the outcome is survival/mortality. The proportion of deaths 28 days after inclusion is a summary measure of the endpoint/variable and not the measure of the outcome. Can you clarify how the definitions align with ICH E9(R1) and the estimand attributes, e.g. the effect measure is the summary measure under ICH E9(R1). A mapping of how the definitions of this guideline align to ICH E9(R1) would be helpful.	Already addressed issue.
S.Walleser Autiero Medtronic	6	Section 2.1.	Organisational impact can be a relevant outcome for some member states in particular for medical devices, and is currently not considered. We suggest that this be included as a potential outcome.	We have decided to focus on patient-centred outcomes as this is the core of HTA.
EFSPI	6	126	The draft guideline advises that it is not desirable if member states request specific effect measures, arguing that "the choice of an effect measure is highly dependent on the underlying assumptions regarding statistical analyses". It should be the health policy decision problem relevant to the member state guiding the choice of effect measure, not the other way around. For example, risk differences may be needed for decision making and health economic modelling, even though a risk ratio or odds ratios may be more readily estimated. However, it is often possible to ensure both a good model fit and the availability of relevant effect measures for decision making (Spiegelman et al, Am J Public Health 2018, doi: 10.2105/AJPH.2017.304105). **Current wording**: "[...] we would advise that specifying an effect measure is not desirable. Indeed, the choice of an effect measure is highly dependent on	This sentence was not meant to discuss in depth the preferences between ratios and differences (which are indeed complementary), but rather pointing out that the interpretation of some effect measure is only valid if certain underlying assumptions are met. The fact they are met or not in the study submitted as evidence by the HTD cannot be known when drafting PICOs. This is why we have made this recommendation.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			underlying assumptions regarding statistical analyses. For example, hazard ratios estimated using a Cox model require that the proportional hazards assumption approximately holds. If not, hazard ratios are not valid estimates and another effect measure should be used, such as the restricted mean survival time. Therefore, it is first the responsibility of the HTD to provide results expressed in terms of effect measures according to good clinical and statistical practice." **Proposed wording**: "It may not always be possible to capture a specific effect measure directly in a statistical model if the model is not appropriate for the data at hand. It is the responsibility of the HTD to perform statistical modelling according to good statistical and clinical practice, ensuring that the model is appropriate for the data at hand. Where possible, coefficients can be transformed to the effect measures requested (e.g. transforming results from a multiplicative model to obtain risk differences etc)"	
MTE	6	126	Definition Outcomes: … treatment and care effectiveness …. Health Technologies,, especially in the field of medical technologies are not limited to treatment but all to enhance the effectiveness of care delivery ensure high value care.	As mentioned above in the guideline, for simplicity, treatment is used in the guideline as one of the many synonyms that can be used to talk about the health technology being assessed.
European Huntington Association	6	127	After symptoms and safety include DISEASE PROGRESSION. There are several diseases where a slower progression would be very beneficial. However, symptoms etc would not change (just slowed).	This will be considered in the next version of the guideline.
Prof. Matthias P. Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer	6	127	Original wording: “[…] treatment effectiveness, such as mortality, remission, health-related quality of life (HRQoL), symptoms and safety.” Comment: It is common practice to categorize treatment effect into mortality, morbidity/symptoms, quality of life and safety. Adding the category of remission as a particularly relevant treatment effect in numerous indications appears reasonable. On top of that, the relevance of achieving a remission should be acknowledged during the entirety of the HTA process, and not only	Thank you for this comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
nat, Laura Könenkamp, Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzens, M.Sc., Steven Krüger, M.Sc., SKC Beratungsgesellschaft mbH			as a definition. We therefore recommend further emphasizing this category.	
EHA	6	127	We suggest using “response to treatment” instead of “remission” because response is a broader term and includes, for instance, also the quantification of measurable residual disease (MRD), which is increasingly used to assess the depth of the response.	This sentence is only introductory and does not consist in an exhaustive list of all the outcomes that can be assessed.
European Huntington Association	6	139	Line 139 states 'Thus, effect measures are primarily understood as a comparison of the measure of outcomes between two interventions groups.' In many disease areas there are no excisting interventions yet. In case placebo is meant as intervention, please clarify.	This will be clarified.
Tanja	6	140	Current text:	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Podkonjak, Takeda			*"Thus, effect measures are primarily understood as a comparison of the measure of outcomes between two interventions groups."* Suggested wording: *"Thus, effect measures are primarily understood as a comparison of the measure of outcomes between an intervention and control group (placebo or different intervention)."* Rationale: Comparisons might not only be made between two interventions groups, but also between an intervention and a "no intervention" group (i.e., placebo).	
Prof. Matthias P. Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer nat, Laura Könenkamp, Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzens,	6	142	Original wording: "However, other statistics can be used to express other aspects of a treatment effect such as the absolute effect or a within-group change." Comment: It appears reasonable to not strictly limit the potential statistical measures and allow for flexibility, especially in cases where it is required.	Thank you for this comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
M.Sc., Steven Krüger, M.Sc., SKC Beratungsgesellschaft mbH				
EFSPI	6	148	EFSPI proposes updating the reference and definition of clinician-reported outcomes to include the ISPOR Good Practice Task Force citation which describes three types of ClinRO assessments (readings, ratings, and clinician global assessments) and describes emerging good measurement practices in their development and evaluation. **Current wording**: "Clinically reported outcomes are assessed by healthcare professionals during clinical examination of a patient and involve clinical judgments of patients' observable signs, behaviours or other physical manifestations." **Proposed wording**: "Clinically reported outcomes are assessed by healthcare professionals during clinical examination of a patient and involve clinical judgments of patients' observable signs, behaviours or other physical manifestations. Clinically reported outcome assessment is conducted and reported by a trained health care professional and requires specialized professional training to evaluate the patient's health status." Reference: Powers JH 3rd, Patrick DL, Walton MK, Marquis P, Cano S, Hobart J, Isaac M, Vamvakas S, Slagle A, Molsen E, Burke LB. Clinician-Reported Outcome Assessments of Treatment Benefit: Report of the ISPOR Clinical Outcome Assessment Emerging Good Practices Task Force. Value Health. 2017 Jan;20(1):2-14. doi: 10.1016/j.jval.2016.11.005. Epub 2017 Jan 10. PMID: 28212963; PMCID: PMC5379997.	It will be considered for the next version of the draft.
Roche	6	148	Roche proposes updating the reference and definition of clinician-reported	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			outcomes (ClinRO) to include the ISPOR Good Practice Task Force citation which describes three types of ClinRO assessments (readings, ratings, and clinician global assessments) and describes emerging good measurement practices in their development and evaluation: "**A ClinRO assessment is conducted and reported by a trained health care professional and requires specialised professional training to evaluate the patient's health status."** *Reference: Powers JH 3rd, Patrick DL, Walton MK, Marquis P, Cano S, Hobart J, Isaac M, Vamvakas S, Slagle A, Molsen E, Burke LB. Clinician-Reported Outcome Assessments of Treatment Benefit: Report of the ISPOR Clinical Outcome Assessment Emerging Good Practices Task Force. Value Health. 2017 Jan;20(1):2-14. doi: 10.1016/j.jval.2016.11.005. Epub 2017 Jan 10. PMID: 28212963; PMCID: PMC5379997.*	
HTAi PCIG	6	Line 153	The document states: "…*the main source of information can be the patients. Patient-reported outcomes (PROs) are defined as "any report of the status of the patient's health condition that comes directly from the patient …"* Whilst we accept the importance of PRO and PROMs, this line reduces any patient input solely to that generated from PROs. The consultation document (line 458) outlines validity and reliability as key, but we know that many PROs, whilst "validated" fail to capture important domains of a condition and its impact and that it is important to seek patient input when contextualising PRO findings and highlighting any gaps. A topical example is the ability of existing PRO questionnaires/PROMs to measure the side-effects of novel immunotherapies (such as checkpoint inhibitors, CAR-T, bi-specifics etc) - the patient perspective is required to plug the data gaps and understanding. If, as the document says, *"… the main source of information can be patients…*", then why does it restrict this to a very narrow set of inputs from PROs? This section should be expanded with an acknowledgement that patient involvement and patient input will be needed to provide further context on the patient experience and to provide views on the appropriateness and relevance of any PRO used to capture patient	Clarifications will be made in the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			experiences in the studies. We also know that many clinical trials still do not measure PROs, so this type of data can be patchy, inconsistent, and missing from industry submissions to HTA bodies. This also increases the importance of the direct patient testaments/involvement and also to the importance of the HTA bodies, through early dialogues, facilitating or even mandating the optimum selection of PRO measures (again with the involvement of patients) as early on as possible so we have optimised data. There is an opportunity here to improve the measurement of PRO in trials, which needs to be optimised. Possible alternative text: *"Similar to clinician reported outcomes, there are several categories of patient reported outcomes: Patient self-reported outcomes can include outcomes measured through patient reported outcomes measures or unstructured reports of the patient experience by representative patients or patient organizations. Technologically assessed patient reported outcomes require the use of technology such as wearables or recording tools managed by the patients."*	
Mihai Rotaru, EFPIA	6	157	Proposed change: "The PRO concept is sometimes equated to HRQoL, **although this is a limited interpretation, as HRQoL** is only a …" Rationale: Reinforces the point that PROs are not just HRQoL	This will be considered for the next version of the draft.
EORTC	6	Line 157	PRO cannot really be equated to HRQoL even if both aim at assessing patient perceptions. At the time of static and dynamic instruments, more specific appreciation of for example physical functioning versus global QoL can have different values and complementarity	This is line with what we say in the guideline.
EFSPI	6	159	**Current wording**: "EQ-5D instrument measures health status as a combination of five broad concepts" **Proposed wording**: "EQ-5D instrument measures health status via five broad concepts."	We think the current sentence is correct as it is as the result is a profile of five responses and not an overall score.
Roche	6	159	Roche proposes rewording "EQ-5D instrument measures health status as a combination of five broad concepts", suggested editing to "EQ-5D instrument	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			measures health status v**ia five broad concepts**"	
Mihai Rotaru, EFPIA	6	161	EFPIA considers observer-reported outcomes (ObsROs) to constitute another main, and separate, source of information. This is of particular relevance in the context of paediatric studies where the patient may be too young to self-report and in the context of neurodevelopmental or neurodegenerative disease areas where the patient may be cognitively impaired, making a self-report unreliable. EFPIA therefore proposes a new paragraph to start after the definition of PRO is complete (line 161) and the definition provided in the EMA Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man can be used. Suggested wording: "***Another source of information can be observers. An observer-reported outcome (ObsRO) is a measurement based on an observation by someone other than the patient or a health professional. This may be a parent, spouse, or other non-clinical caregiver who is in a position to regularly observe and report on a specific aspect of the patient's health. An ObsRO measure does not include medical judgment or interpretation. Generally, ObsROs are reported by a parent, caregiver, or someone who observes the patient in daily life. For patients who cannot respond for themselves (e.g., infants or cognitively impaired), we encourage observer reports that include only those events or behaviours that can be observed. As an example, observers cannot validly report an infant's pain intensity (a symptom) but can report infant behaviour thought to be caused by pain (e.g., crying).***"[1] EFPIA proposes that a distinction should be made between an ObsRO and a proxy-reported measure. As currently written, the two concepts are mixed-up, but those are two distinct concepts. A proxy is someone who acts on behalf of the patient to complete a PROM. For example, if a patient's disease worsens during the trial, and the patient does not have enough strength to write the answers, a nurse could ask the questions of the PROM out loud, and document the answer given by the patient. EFPIA proposes the guideline be revised to clearly distinguish "observer-reported	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			outcome" and "proxy-reported outcome" and suggests referring to the EMA Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man: *"A distinction should be made between an ObsRO and a proxy-reported outcome. A proxy-reported outcome is a measurement based on a report by someone other than the patient reporting as if he or she is the patient. A proxy-reported outcome is not a PRO. A proxy report also is different from an observer report where the observer (e.g., clinician or caregiver), in addition to reporting his or her observation, may interpret or give an opinion based on the observation".*[1] Reference: 1. EMA - Appendix 2 to the guidance on the evaluation of anticancer medicinal products in man, 1 April 2016. Accessed 6th October 2022: https://www.ema.europa.eu/en/documents/other/appendix-2-guideline-evaluation-anticancer-medicinal-products-man_en.pdf	
Roche	6	161	Roche considers observer-reported outcomes (ObsROs) to constitute an additional main, and separate, source of information. This is of particular relevance in the context of paediatric studies where the patient may be too young to self report and in the context of neurodevelopmental or neurodegenerative disease areas where the patient may be cognitively impaired, hence making self report unreliable. Roche proposes a new paragraph to start after the definition of PRO is complete (line 161) and the definition provided in the EMA Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man can be used: "**Another source of information can be observers. An observer-reported outcome (ObsRO) is a measurement based on an observation by someone other than the patient or a health professional. This may be a parent, spouse, or other non-clinical caregiver who is in a position to regularly observe and report on a specific aspect of the patient's health. An ObsRO measure does not include medical judgement or interpretation. Generally, ObsROs are reported by a parent, caregiver, or someone who observes**	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			the patient in daily life. For patients who cannot respond for themselves (e.g., infants or cognitively impaired), we encourage observer reports that include only those events or behaviours that can be observed. As an example, observers cannot validly report an infant's pain intensity (a symptom) but can report infant behaviour thought to be caused by pain (e.g., crying)." Moreover, Roche proposes that a distinction should be made between an ObsRO and a proxy-reported measure. Roche proposes referring to the EMA Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man: **"A distinction should be made between an ObsRO and a proxy-reported outcome. A proxy-reported outcome is a measurement based on a report by someone other than the patient reporting as if he or she is the patient. A proxy-reported outcome is not a PRO. A proxy report also is different from an observer report where the observer (e.g., clinician or caregiver), in addition to reporting his or her observation, may interpret or give an opinion based on the observation"**. *Reference: European Medicines Agency. Appendix 2 to the guidance on the evaluation of anticancer medicinal products in man, 1 April 2016. Accessed 6th October 2022: https://www.ema.europa.eu/en/documents/other/appendix-2-guideline-evaluation-anticancer-medicinal-products-man_en.pdf*	
Tanja Podkonjak, Takeda	6	164	Takeda recommends distinguishing a proxy reported outcome from observer-reported outcomes. Suggested addition at the end of the paragraph: Proxy report is to be distinguished from observer-reported outcome (ObsRO), which is a measurement based on a report of observable signs, events or behaviours related to a patients' health condition by someone other than the patient or health professional. Generally, ObsROs are reported by a parent, caregiver, or someone who observes the patient in daily life.[1]	Clarifications will be made in the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Reference: BEST glossary https://www.ncbi.nlm.nih.gov/books/NBK338448/#IX-C	
M. Ermisch – GKV-SV	6	166 ff	The information given on the use of patient-generated health data needs further contextualisation. The excitement for data generated by e.g. wearables must not result in neglecting the need for proper and a priori validation of measurement and results by HTD.	The paragraph will be modified for the next version of the draft.
Mihai Rotaru, EFPIA	7	178 -219	In Section 2.2, the suggested language to MS in completing a PICO survey recommends using statements of preference - "measured preferably". This is seemingly contradicted in Section 3.3, where "for all outcomes requested in the assessment scope, the HTD should provide data". Is there an intention to treat surrogate outcomes inconsistently in this regard? Please clarify.	It will be clarified for the next version of the draft.
EFSPI	7	178-210	The wording for the scoping process on the outcomes should be accompanied by justification/rationale to enable HTD to better understand the request. These should also be aligned to the study design, otherwise it becomes an immediate disadvantage to HTD for not being able to provide them. Furthermore, they may violate statistical principles and/or being a data dredging exercise.	Details on how the scoping process will be conducted are available in the guideline scoping process. There is the possibility of JSCs to allow discussions between HTDs and HTAbs for a better alignment of HTDs research questions and HTAbs research questions. Assessment scope will be blinded which limits the risk of data dredging. Other guidelines, such as the D4.5 and D4.6 will allow assessor to report the elements of methods and results that will allow HTAbs to perform adequate decision-making.
Mihai Rotaru, EFPIA	7	183 - 208	The fulfilment of a request depends strongly on the clinical trial(s) which have been submitted to and approved by regulators. Therefore - as stated in previous consultations - the evidence from pivotal trials, measured by pre-defined and approved endpoints submitted to EMA need to be taken into account in the scope. Understanding these endpoints will also ensure Member States can avoid unnecessary requests which cannot be fulfilled. Involvement of the HTD in the scoping process can help provide further insight to the evidence available (including	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			timepoint assessments) which would make the process more efficient, relevant and allow the assessors to consolidate the PICOs further when multiple outcome measures are requested for an outcome.	
EFSPI	7	183-198	The draft guideline encourages that research questions include neither specific outcome measures, timing, or effect measures. While this is mentioned only as a preference, it is a scientific concern that there seems to be more focus on maximizing the ability to achieve a result, than addressing clearly formulated decision problems. Specifically, the view here is very different from that taken in the ICH E9 estimand addendum, which explicates the importance of being clear about the specific endpoint (outcome measure), timing, and summary measure. A better way to maximize the ability to achieve a result would be to ensure that the HTD is allowed timely input to the scoping discussion, to align on outcome measures that realistically can be reported.	JSCs will be an opportunity to better align clinical development with HTAbs needs. Regarding the involvement of HTDs during the scoping process, this is out of the scope of this guideline.
Natacha Bolanos, Lymphoma Coalition	7	183-198	In the absence of standard measures of HRQL, and despite manufacturers investing effort to address inconsistencies between measures, there is no guidance on how output from different measures can be usefully streamlined. The harmonization process would be greatly simplified by all EU Member States recommending a single instrument to measure HRQL. However, a wide range of instruments are currently in use, and in evolving health care landscapes their relevance may be temporary.	It is not reasonable to assume only one HRQoL instrument can cover all the variety of the research questions that can be addressed in the medical field.
Roche	7	199-208	Current wording: "While this practical guideline does not endorse any criteria to be filled by MS when requesting health outcomes, we would advise that specifying an effect measure is not desirable. Indeed, the choice of an effect measure is highly dependent on underlying assumptions regarding statistical analyses. For example, hazard ratios estimated using a Cox model require that the proportional hazards assumption approximately holds. If not, hazard ratios are not valid estimates and another effect measure should be used, such as the restricted mean survival time. Therefore, it is first the responsibility of the HTD to provide results expressed in terms of effect	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			measures according to good clinical and statistical practice. Nonetheless, if an MS wants to specify an effect measure, this should be done using the previously mentioned template: "[Outcome of interest] with treatment effect expressed preferably as [insert effect measure]"." Proposed wording: ""While this practical guideline does not endorse any criteria to be filled by MS when requesting health outcomes, we would advise that specifying an effect measure is not desirable. Indeed, the choice of an effect measure is highly dependent on underlying assumptions regarding statistical analyses. For example, hazard ratios estimated using a Cox model require that the proportional hazards assumption approximately holds. If not, hazard ratios are not valid estimates and another effect measure should be used, such as the restricted mean survival time. Therefore, it is first the responsibility of the HTD to provide results expressed in terms of effect measures according to good clinical and statistical practice. ~~**Nonetheless, if an MS wants to specify an effect measure, this should be done using the previously mentioned template: "[Outcome of interest] with treatment effect expressed preferably as [insert effect measure]".**~~ " (deletion of last sentence) Rationale: The Guideline should aim at internal consistency and, therefore, avoid conflicting recommendations. We suggest dropping the last sentence, which is not consistent with the rest of the paragraph.	
Mihai Rotaru, EFPIA	7	192-198	We suggest adding the following wording to this section: *"It is the responsibility of the MS to justify the choice of timing specified. The JCA should only include results from such outcomes if sufficiently justified".* Rationale: Data for a specified time point may be inappropriate, e.g. because it is intrusive to the patients, their families or healthcare system to obtain data beyond the clinical trial follow-up. If a MS may want to specify the timing of outcome assessment, in addition to the suggested template, the MS should justify why the specified timing	In accordance to the HTAR, MS do not have to justify their outcomes request.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			would be considered appropriate.	
Roche	7	196-202	Roche notes that in line 202 hazard ratios are included as an effect measure. However, in line 196 mixed models are mentioned without any consideration that these provide an effect measure or a statistical test, both of which rely on assumptions that would need to be assessed. Roche proposes to add to line 206: **"Many statistical techniques, including mixed models, rely on assumptions that require assessment."**	The main point of this paragraph is just to illustrate that effect measures are intertwined with the underlying assumptions on which they rely, therefore the general recommendations that during the assessment scope effect measures should not be specified by HTAbs. It is common knowledge that statistical analyses rely on assumptions.
Sebastian Werner vfa	7	193-198	*"A request such as "rate of major adverse cardiovascular events 2 years after inclusion" specifies a timing, but also at the risk of not obtaining results, if, for example, follow-up was not sufficiently long in the clinical study submitted as evidence. Such a request of one specific time point could also hamper the presentation of results according to statistical modelling such as mixed models for repeated longitudinal data. A general recommendation could also be to formulate a request as such: "[Outcome of interest] measured preferably at [insert timing of assessment]"."* The draft guideline addressed concerns if the follow-up was considered not sufficiently long in the clinical study submitted as evidence and recommended to formulate a request. Preferences may vary between Member States. One MS might be interested in ""rate of major adverse cardiovascular events 2 years after inclusions" another MS is interested in 12 months or 18 months, 3 years after inclusion. On the other hand, the data collection of a study usually covers the clinically most important periods, where changes in symptoms or side effects are expected. In addition, data collection period and frequency of assessments also considers the factors that would compromise a scientifically sound evaluation e.g., when the rate of missingness is considered too high to draw reasonable conclusion and acceptable level of patient burden. Therefore, the request for a specific time point should be harmonized across Member States and adequately reflect the clinical and technological context and the evidence generated by the HTD. To inform and align Member states	Interactions between HTDs and HTAbs are covered in other guidelines.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			about the possible and useful time points, the *Scoping Process* must consider the HTD's clinical evidence (incl. SP, SAP) to ensure that meaningful time points of the study are requested by Member States. Further, requests for specific time points should be also clarified as part of *Joint Scientific Consultations* that are carried out in parallel with the EMA-Advice to ensure that meaningful time points are defined that provide necessary evidence for EU-HTA and regulatory approval. Secondly, as marketing authorisation could be based on a positive benefit-risk evaluation using interim results of a study, only sparse data might be available to provide a reasonable precise estimate e.g. for the rate of major adverse cardiovascular event 2 years after inclusion. Statistical methods could be used to project or simulate certain outcomes for a specific outcome at a certain time point accordingly. Such methods could be useful to decrease uncertainty and thus should be considered for JCA. The vfa recommends adding the wording: *"To align on meaningful time points Member States should consider the HTD's clinical evidence at time of the Scoping Process and Member States should clarify their needs in Joint Scientific consultations prior to study initiation. Statistical methods to estimate the effect at certain time points using modelling approaches can be applied. "*	
EFSPI	7	195-198	The guidance addressed concerns if the follow-up was considered not sufficiently long in the clinical study submitted as evidence and recommended to formulate a request "[Outcome of interest] measured preferably at [insert timing of assessment]". Preferences may vary between MS. One MS might be interested in "rate of major adverse cardiovascular events 2 years after inclusions" another MS is interested in 12 months or 18 months, 3 years after inclusion. On the other hand the data collection of a study usually covers the clinically most important periods, where changes in symptoms or side effects are expected. In addition, data collection period and frequency of assessments also considers the factors that would compromise a scientifically sound evaluation e.g. when the rate of missingness is considered too high to draw reasonable conclusion and also acceptable level of patient burden.	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			In essence the request for a timing of assessment should be harmonized across member states, it should adequately reflect the disease setting and clinical context. So a joint scientific advise meeting (including REG and HTA) is recommended to clarify the needs so that the study can be designed accordingly. Second, as marketing authorization could be based on a positive benefit-risk evaluation using interim results of a study, only sparse data might be available to provide a reasonable precise estimate e.g. for the rate of major adverse cardiovascular event 2 years after inclusion. Statistical methods could be used to project or simulate certain outcomes for a specific outcome at a certain time point accordingly. Such methods could be useful to decrease uncertainty and thus should be taken into account for the added benefit assessment. In summary please consider to change the wording from: **Current wording**: “A general recommendation could also be to formulate a request as such: “[Outcome of interest] measured preferably at [insert timing of assessment]”. **Proposed wording**: “A general recommendation could be to align key time point of interest prior to study initiation in a joint scientific advise meeting, or to pre-specify statistical methods to estimate the effect at certain time points also using modelling approaches.”	
EFSPI	7	203-206	**Current wording**: "For example, hazard ratios estimated using a Cox model require that the proportional hazards assumption approximately holds. If not, hazard ratios are not valid estimates and another effect measure should be used, such as the restricted mean survival time. Therefore, it is first the responsibility of the HTD to provide results expressed in terms of effect measures according to good clinical and statistical practice" While we can appreciate the overall point being made here, EFSPI thinks this example runs the risk of EUnetHTA implying that if the proportional assumption do not hold, it is ok to avoid presenting hazard ratios results. This	This sentence will be deleted as the main point of the paragraph is already clear.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			may be at odds with the expectation that if the hazard ratio is pre-defined as an analysis it should not be disregarded on the basis of an assessment of PH being deemed as violated (especially since proportional hazard assessment is somewhat subjective) We would recommend to remove this example, as the point the guidance is making is still sufficiently clear without it.	
Mihai Rotaru, EFPIA	7	175-177	Proposed change: *"For example, the Disease Activity Score (DAS 28) for rheumatoid arthritis* ***which*** *requires clinical, technological and patient-reported elements [13]* ***or the Mayo Score (Schroeder et al. 1987)****.* Rationale: In addition to DAS, the Mayo Score is another hybrid measure that can be added. The Mayo Score was developed as a composite disease activity index for use in clinical trials. The original description of the Mayo Score included an assessment of 2 patient-reported outcomes [PROs; stool frequency (SF) and rectal bleeding (RB)], the endoscopic appearance of the mucosa (endoscopic score, ES), and a Physician's Global Assessment (PGA), each of which were scored on a scale from 0 to 3, giving a maximum total score of 12. Reference: Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. N Engl J Med. 1987;317:1625–1629.	The use of the DAS-28 is only here to illustrate a broader issue, not to validate an exclusive list of appropriate scores.
Mihai Rotaru, EFPIA	7	179 - 181	Current wording: *"During the assessment scoping stage for JCA, the definition of outcomes requested by MS should be as appropriate as possible, as this can impact assessment of the results submitted by a HTD in a JCA report."* EFPIA notes that the same statement could be made for populations, sub-populations, sub-groups and comparators. We therefore recommend removing this statement. If it is maintained, the reader would benefit from defining and clarifying what 'appropriate as possible' is intended to mean, and who defines it.	This will be clarified for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Mihai Rotaru, EFPIA	7	184 - 186	Proposed change: ~~However, the HTD could provide a result using a measure of the outcome that could be considered inappropriate (e.g., because the measure is appraised as having an insufficient level of validity).~~ (delete sentence) Rationale: It is not the purpose of the scoping to determine the validity of an outcome or outcome measure, as that should be undertaken as part of the JCA itself as outlined by the Regulation. EFPIA proposes this statement is removed. Furthermore, the appraisal of outcomes and associated certainty and validity is a Member State responsibility, and this guideline should not discuss appraisal. EFPIA would like to reiterate that the totality of evidence should be considered in an JCA, and not preclude relevant information on a new technology.	The purpose of this sentence will be clarified for the next version of the draft.
Mihai Rotaru, EFPIA	7	203-205	Proposed change: *"If not, hazard ratios are not valid estimates and another effect measure should be used, such as the* ***difference in*** *restricted mean survival time."* Rationale: HR is a relative measure between two treatment arms. The equivalent with RMST methodology is either the difference or the ratio in RMST.	This sentence will be deleted as the main point of the paragraph is clear without it.
Matias Olsen, EUCOPE	7	172-174	The use of digital technologies is now widespread, we suggest clarifying that use of such technologies due to lack of sufficient digital literacy is primarily a challenge in certain age groups.	We do not think age is the only factor that can influence digital literacy.
Bayer	7	179-181	Statement: During the assessment scoping stage for JCA, the definition of outcomes requested by MS should be as appropriate as possible, as this can impact assessment of the results submitted by a HTD in a JCA report. Proposal: Please replace "as appropriate as possible" with "accepted by medical, scientific, and regulatory standards". Furthermore, please add a general consideration that MS cannot reject outcomes that are otherwise	The message of this paragraph will be clarified for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			accepted by medical, scientific, and regulatory standards.	
Bayer	7	184-186	Comment: It is stated that HTD could provide a result using a measure of the outcome that could be considered inappropriate. However, this should be unlikely to occur, if the outcome and measure have been accepted for the regulatory approval process.	Validity is not a on/off concept. It comes in degrees. Moreover, a COA can be fit for purpose for one inference and not for another. Therefore, it cannot be expected that MS will appraise any result on an outcome as a fit for purpose one.
EFSPI	7	178-179	**Current wording**: "During the assessment scoping stage for JCA, the definition of outcomes requested by MS should be **as appropriate as possible**, as this can impact assessment of the results submitted by a HTD in a JCA report." We would welcome a clearer explanation as to exactly what is meant about outcomes being defined 'as appropriately as possible'.	It will be clarified for the next version of the draft.
EFSPI	7	191-192	General considerations refer to the definition of an outcome, in particular specifying an appropriate PRO endpoint. The guideline says “To alleviate this issue, a general recommendation could be to formulate a request as such: “[Outcome of interest] measured preferably as [insert measure]”.” The recommendation should be extended and reference given to the SISAQOL-IMI2 initiative aiming for setting International Standards in Analysing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials. Similarly to the reference to COMET for COS funded also by IMI2. **Proposed wording**: “To alleviate this issue, a general recommendation could be to formulate a request as such: “[Outcome of interest] measured preferably as [insert measure]” and to consider international standards in analysing PROs and Quality of Life Endpoints in cancer clinical trials as provided by SISAQOL-IMI.”	This guideline is not only about cancer trials. We do not think it is useful to endorse one restricted initiative.
Sebastian Wernervfa	7 9	191-192 252-253	*“To alleviate this issue, a general recommendation could be to formulate a request as such: “[Outcome of interest] measured preferably as [insert measure]”.”* *“Efforts are being conducted to identify a standardised set of outcomes that*	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			should be measured and reported, as a minimum, in all clinical trials in specific areas of health or healthcare, defined as a core outcome set (COS)" The recommendations should include, and reference given to the SISAQOL-IMI2 initiative aiming for setting International Standards in Analysing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials. Similarly, to the reference to COMET for COS funded also by IMI2.	
Bayer	7	201-202	Original statement: "Indeed, the choice of an effect measure is highly dependent on underlying assumptions regarding statistical analyses". Comment: Even though it is not included in PICO, the summary effect measure is an important component of the research question. As a result, it is included in the estimand definition ("population-level summary measure") issued by the ICH E9 (R1) Addendum, recognized by FDA and EMA in the regulatory setting. The estimands framework demands that estimands (i.e., research questions) are articulated prior to statistical analysis. The statements in this paragraph suggest that the estimand is dependent on the statistical methodology and its modelling assumptions. This contradicts guidance in the ICH E9 (R1) Addendum. Note that specifying a summary measure is also important in HTA, e.g., for defining effect modification, necessary to assess generalizability and external validity.	These considerations will be clarified for the next version of the draft.
AIM – International Association of Mutual Benefit Societies	7	2.2	The formulation *"[Outcome of interest] measured preferably as [insert measure]"* is not prescriptive enough and still gives thee opportunity for HTDs to not follow the preferable option. Couldn't we instead say "[Outcome of interest] measured ~~preferably~~ as [insert measure] OR [insert measure] OR [insert measure]" ?	This formulation is here to allow a reasonable compromise allowing MS to indicate what they would prefer why allowing HTDs to provide a result on the corresponding outcome if another measure was used.
AIM – International Association of	7	2.2	In general, the guideline does not say what shall happen if the HTD does not follow the recommendations on how to express outcomes, be it in measure, assessment timing, or effect measure	This will be factually reflected in the JCA report and MS will appraise the results the way they see fit.

EUnetHTA 21 Public Consultation Comments and Responses
Of D4.4 - Endpoints

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Mutual Benefit Societies				
M. Ermisch – GKV-SV	7	Sec. 2.2	We highly appreciate the fact that this section openly discusses the possible problem arising from mismatches between the PICO defined during JSC and scoping on the one hand and data supplied by the HTD in the HTA-dossier. HTA cannot disallow HTD the use of different scales and methods for the determination of benefits. Thus, one way forward might indeed be, to define endpoints and endpoint categories relevant for national decision-making and give recommendation on scales that are usually regarded to be relevant and reliable. If the HTD decides to change these or use different ones, proof of validity is within its responsibility.	Thank you for this comment.
GSK	7	Section 2.2	• Please clarify the importance of HTDs having the ability to receive input from MS on PICO requirements when clinical trials are designed, e.g. through scientific advice. If there are specific outcomes of interest for MS that are not considered when designing the trial, it may not be feasible or scientifically possible for the HTD to derive estimates of treatment effectiveness for these outcomes. 1- Please cross-reference other JCA methodology guidelines on how to define outcomes that can be estimated with minimal statistical bias and how alignment on outcomes can be achieved through the scoping process, in particular when designing clinical trials.	JSCs and interactions between HTDs and HTAbs are the concerns of other guidelines.
Norbert Gerbsch for IGES Institut GmbH and Health Econ AG	7 8	2.2 General considerations 199-209 Points of attention for the assessment	**Comment:** *"...we would advise that specifiying an effect measure is not desirable."* *„Effect measures should not be specified by MS. The HTD is responsible for presenting results using appropriate effect measures in accordance with good clinical and statistical practice."* It is welcomed that the guideline explicitly enables MS to specify effect measures i.e. by "[Outcome of interest] measured preferably as [insert measure]". Nevertheless besides opening this option it is at the same time termed as undesirable.	As the assessment scope is performed after clinical development, these recommendations were made to find compromises accommodating this situation. In addition, JSCs will be the place for fostering a better alignment between clinical development and MS needs.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
		scoping process	How should the HTD gain information on outcomes of relevance for the JCA if no respective specification is given i.e. the MS? Should not the PICO determination provide a valid and reliable guideline for the HTD? It is therefore suggested to encourage the specification of effect measures by MS or – at least – to adopt a neutral position in this respect.	
EFSPI	7	181	"Therefore, general guidance…" Please include the references to the guidance that is being considered or link to the citation for the existing EUnetHTA guidelines on endpoints.	This will be clarified for the next version of the draft.
Roche	7	181	Roche requests clarification regarding the "general guidance" referred to as useful for formulating outcomes. Please clarify if this general guidance is provided directly in section 2.2 or provide a citation if it is found elsewhere.	Duplicated comment.
Mihai Rotaru, EFPIA	7	183	"Defining an outcome at the broadest level (e.g., HRQoL without further specifications) maximises the opportunity to obtain a result." Comment: We disagree that defining an outcome broadly maximises the opportunity to obtain a result. For example, a general HRQOL could take longer to change than perhaps a disease-specific symptom. The intention of the paragraph seems to be that it is important to define an outcome more precisely, so this sentence should reflect that purpose as it could be potentially misleading as it is currently written.	The issue at hand was not generic versus specific HRQoL. This will be clarified for the next version of the draft.
EFSPI	7	186	EFSPI proposes clarifying what is meant by 'adequacy' in line 186. For example, this could include cross-referencing the later validity and reliability section. **Current wording**: "The adequacy of the measure of the outcome provided by the HTD therefore needs to 186 be appraised by the MS on the basis of the elements reported within the JCA."	This will be clarified for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Proposed wording: "The adequacy (see section 5 for an overview of validity and reliability measurement properties) of the measure of the outcome provided by the HTD therefore needs to be appraised by the MS on the basis of the elements reported within the JCA."	
Roche	7	186	Roche proposes clarifying what is meant by 'adequacy' in line 186. For example, this could include cross-referencing the later validity and reliability section. The guideline should seek to establish a pan-European consensus on the approaches and criteria that are acceptable to establish the scientific adequacy of an outcome measure. "The adequacy **(see section 5 for an overview of validity, reliability and interpretability of scales)** of the measure of the outcome provided by the HTD therefore needs to be appraised by the MS on the basis of the elements reported within the JCA."	Duplicated comment.
Liebenhoff, BAH	7	192	"[Outcome of interest] measured preferably as [insert measure]" In certain circumstances there would be several options of "insert measure" Therefore, please add "list of …"	We did not write this recommendation as being restricted to one measure only.
Bayer	7	192, 198, 208	Comment: How binding would the requests for preferred measures and timepoint be interpreted? There is no clearly statement on this and it leaves room for misinterpretations.	This is out of the scope of the guideline.
S.Walleser Autiero Medtronic	8	Summary	We welcome the suggestion that effect measures should not be specified by the MS but that this is subject to HTD's judgement on what the right effect measures should be for the specific technology and condition under evaluation.	Thank you for this comment.
Roche	8	215/3.1	Current wording: "This may be the case for surrogate outcomes and biomarkers (see the definitions in Section 3.2)" Proposed wording: "This may be the case for surrogate outcomes ~~**and biomarkers**~~ (see the definitions in Section 3.2)" Rationale: surrogate outcomes may include clinical outcomes or biomarkers	It will be corrected for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			and we consider that there is no need to call out biomarkers separately.	
Bayer	8	Summary box	Comment: Outcomes and measures for MS to specify during the scoping process basically will be confined to those of the pivotal study – for clinically reported outcomes, PRO and safety outcomes. HTDs will not be able to produce or alter outcomes and measures beyond those defined by the pivotal study. This should be made clear in the deliverable.	We do not think such clarification is not useful as it is obvious.
Tanja Podkonjak, Takeda	8	Summary Box	Summary (bullet 3) Current wording: *"The measure of an outcome defines accurately how the outcome is assessed. Effect measure are primarily statistics used to compare the measure of outcomes between two intervention groups."* Suggested rewording: *"The method used to measure an outcome defines the assessment of the outcome. Effect measure are primarily statistics used to compare the measure of outcomes between groups."* Rationale: As currently stated, "accurately" implies a measure of statistical distribution (i.e., precision and accuracy) while the statement is intended to discuss methodology. Furthermore, the groups being compared in these studies are not always limited to two nor always have an intervention group (i.e., a placebo or control may be present).	It will be clarified for the next version of the draft.
Tanja Podkonjak, Takeda	8	Summary Box	Points of attention for the assessment scoping process (bullet 3) Current text: *"If an MS wants to specify a time point for assessment, the wording should follow this template: "[Outcome of interest] measured preferably at [insert timing of assessment]".* This bullet point is seemingly contradictory to the statement on page 8, line 195-196 above: "*Such a request of one specific time point could also hamper*	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			the presentation of results according to statistical modelling such as mixed models for repeated longitudinal data." It is not clear if measurements at specific time points are recommended or under which circumstances, they may be appropriate – we recommend further clarification. Takeda notes the guideline statement that requesting measurements at specific timings may increase 'risk of not obtaining results' and is concerned about the implications of MS requesting specific and possibly different timings of measurement. We suggest the guidance be clearer on requests for timing and if kept in the guidance, suggest any specific timing request be founded in strong clinical rationale.	
Tanja Podkonjak, Takeda	8	Summary Box	Preferable terms requested in scoping: *"[Outcome of interest] measured **preferably** as [insert measure]".* *"[Outcome of interest] measured **preferably** at [insert timing of assessment]".* *"[Outcome of interest] with treatment effect expressed **preferably** as [insert effect measure]".* It is not clear what is the impact on the JCA submission of an HTD not being able to meet the items requested as 'preferable' in scoping requests. We request flexibility be introduced and items, timings or measurements listed as 'preferable' not be considered in the status of completeness of a JCA dossier.	JCA report will be produced based on evidence submitted by HTDs and based on the recommendations made in the different guidelines and MS will appraise the results the way they see fit.
EFSPI	8	212-220	We acknowledge that outcomes supporting the benefit-risk assessment might be "less suitable for the needs of JCA". However, the healthcare and treatment decisions strongly depend on the prescribing information (PI). Hence all outcomes described in the PI should be considered for JCA purposes and used as a common "core" outcome set among all member states. **Current wording**: "Some outcomes may be fully acceptable as support for the risk/benefit ratio assessment of a certain therapy but are less suitable for the needs of JCA." **Proposed wording**: "Some outcomes may be fully acceptable as support for	Thank you for the suggestion. Core set of outcomes is not the same as outcomes listed in the PI.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			the risk/benefit ratio assessment of a certain therapy but are less suitable for the needs of JCA. However, all outcomes described in the prescription information should be considered for JCA purposes." Additionally, **Current wording**: "The acceptability of an outcome is subject to MS interpretation of their relevance within their national process for decision making and thus may differ between MS." **Proposed wording**: "The acceptability of an outcome is subject to MS interpretation of their relevance within their national process for decision making and thus may differ between MS. At least all outcomes described in the prescription information should be considered as a core set of outcomes."	
François Houÿez, Eurordis	8	223-231	When there is not such an overlap in the choice of what are considered patient-centred outcomes for JCA with PICO requests, can a different PICO reflecting the choice of the patients who have been consulted also be submitted to the HTD? A Patient Centred PICO?	Thank you for your comment. This procedure is not mentioned in HTAR. PICO guideline specifies requirements for submission. This suggestion is out of scope for this guideline.
S.Walleser Autiero Medtronic	8	215-222	The important difference between needs for outcomes for regulatory approval (risk/benefit ratio) versus needs for JCA (preferably "final outcomes") is recognized. It should further be clarified under which circumstances exactly surrogate endpoints would be accepted (in particular if the timing of the JCA does not allow the full collection of final endpoints, or if evaluation of final outcomes in an RCT is not feasible) and how real-world evidence can support the determination of "final outcome"/impact of a technology.	Thank you for the comment. Definition of of relevant surrogate endpoints for the JCA is out of scope for this guideline.
Matias Olsen, EUCOPE	8	226-229	To ensure a consistent application, this should not be left to the subjectivity of the assessors. A check list of what is and what is not a patient-centred outcome should be produced.	Thank you for your comment. We will consider it for next edition.
Hayley Chapman,	8	232-235	As per a previous comments, while the document lists "Classifications such as the International Classification of Functioning, Disability and Health of the World 233 Health Organization (WHO) [18], the Wilson and Cleary	Thank you for your comment. We will consider it for next edition.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
PFMD			biopsychosocial model [19] and the Montreal 234 Accord on Patient-Reported Outcomes [6] can provide further information on outcomes that can be 235 assessed in healthcare", the taxonomy of impact as referenced by https://pubmed.ncbi.nlm.nih.gov/29288712/ should also be included.	
Matias Olsen, EUCOPE	8	218-220	This guidance should encourage that outcomes that are used for the joint clinical assessment, and have been selected based on the commonly agreed methodologies at the EU level, should be applied homogenously at Member State level.	Thank you. We hope that the alignment of methodology will take place with time.
GSK	8	213-215	2- Whilst additional endpoints may be required compared to benefit/risk assessment of the regulatory agencies, there should be guidance to ensure that endpoints which are considered acceptable from a regulatory perspective are also accepted by MS for the joint clinical assessment and in their decision-making.	Thank you. MS decide what kind of endpoints are relevant for their national assessments.
GSK	8	218-220	3- Whilst MS may wish to have evidence on specific endpoints, guidance should be provided to ensure that endpoints assessed in the JCA should be accepted by all MS in their decision-making.	Thank you. See above.
Hayley Chapman, PFMD	8	218 -220	The document states "The acceptability of an outcome is subject to MS interpretation of their relevance within their national process for decision making and thus may differ between MS". This is concerning as patients/the patient community should have a greater influence over the relevance of an outcome.	Thank you. Involvement of patients is not a part of this guideline but the patients' inputs are included in the process of HTA.
Liebenhoff, BAH	8	214-215	"This may be the case for surrogate outcomes and biomarkers". Giving such examples in this section is not necessary. Limitations are described in detail later. For that, please delete this sentence.	Thank you
EFSPI	8	226-227	EFSPI proposes that qualitative research is referenced to support the decision on what patient-centered outcome should be included in the PICO question and that collaboration with caregivers (where appropriate) as well as patients is important. The section on validation (section 5) could also be cross-referenced here. **Current wording**: "Deciding what is a patient-centred outcome for the PICO question for a particular therapy should ideally be done in close collaboration	Thank you for your input.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			with patients and healthcare professionals…" **Proposed wording**: "Deciding what is a patient-centered outcome for the PICO question for a particular therapy should ideally be established through a literature review and where required through qualitative research with patients, caregivers (where appropriate) and healthcare professionals…"	
Prof. Matthias P. Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer nat, Laura Könenkamp, Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzens, M.Sc., Steven Krüger, M.Sc., SKC	8	209-210 summary	Original wording: "If an MS wants …" Comment: In view of the HTA it is common to assign the responsibility for "using appropriate effect measures" to the manufacturer, i.e., to the HTD. Here, a further specification allows member states to request specific outcome measures, time points or effect measures on an individual basis. In this regard, two aspects should be highlighted. First, different requirements can lead to a high effort for both sides. Second, experience shows that partially HTA bodies requests (knowingly) cannot be fulfilled by the manufacturer, e.g., when it comes to providing data for a not assessed endpoint. Specific requests by member states should therefore be realistic and feasible, in order to allow the manufacturer to provide solutions. In addition, exceptions should be allowed if a convincing rationale is given.	Thank you. The PICO guidelines specify the requirements.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Beratungsgesellschaft mbH				
S.Walleser Autiero Medtronic	8	226-227	The relevance of collaboration with patients in defining the relevant outcomes for a therapy is emphasized and we want to emphasize the importance of this. We take this opportunity as well to emphasize in this context as previously the need for clear guidance and a process on how patients and patient organisations will be recruited and their feedback collected in the JCA process.	Thank you. Patient involvement in HTA is described in a separate guideline.
Roche	8	225-226	Roche acknowledges the statement that "the term "patient-centred outcomes" refers to outcomes that directly measure mortality, morbidity and outcomes related to patients' feelings, beliefs, preferences, needs and functions (such as the ability to perform activities in daily life)". Due to the variation in measurement, Roche proposes adding the following text: **"these are examples of outcomes that may be relevant to a condition or a treatment approach, and that the measurement of these concepts is dependent on the disease area and hence not all of these may be relevant or appropriate to assess. The JSC provides an opportunity for early discussions around the selection and approach to measurement to ensure harmonization of evidence required at the pan-European level ".**	Thank you for your suggestion. We will consider it.
Mihai Rotaru, EFPIA	8	Section 3	We note that this section, despite being called Clinical Relevance, focuses on patient-centred outcomes and surrogate outcomes (and how they relate to patient-centred outcomes). Patient-centred outcomes are an important aspect of both regulatory and HTA, but the Regulation does not define clinical assessment solely in these terms. We refer EUnetHTA 21 to our General Point 4 which outlines why clinical endpoints themselves are highly relevant for a Joint Clinical Assessment and HTA bodies across the EU. This section should reflect that clinical outcomes themselves are important for HTA and discuss them in addition to their role as surrogate markers.	This issue is already addressed. Thank you.
EORTC	8	Section 3	This is unclear: if the question is properly defined as clinically meaningful, the	Thank you. We agree.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			variability should be lesser. Otherwise, such approach opens door to inequalities across EU citizens. A medically defined priority has no border.	
HTAi PCIG	8	Table, Line 4	.. measure**s** (plural)	Thank you.
Mihai Rotaru, EFPIA	8	Summary Box after line 209	Composite outcomes are not mentioned in the guideline. We request these be included in the guideline as potential outcomes as well. It is suggested to add a bullet: **The relevance of outcome measures validated with Professional Medical Associations.** ***Proposed change 1:*** • *"Effect measure**s** are primarily statistics used to compare the measure of outcomes between ~~two intervention~~ groups. Other statistics can be used for other purposes (absolute effect, within-group change). "* Rationale: The groups being compared in these studies are not always limited to two or always have an intervention group. The measure of outcomes may be between an intervention of interest group and a control group (no intervention or administration of a placebo or control intervention). ***Proposed change 2:*** The existing bullet point six be amended by addition of the underlined text: • "If an MS still wants to specify a time point for assessment ***for which there is sufficient justification***, the wording should follow this template ..." ***Proposed change 3:*** The existing bullet point eight be amended by addition of the underlined text: • "If an MS still wants to specify an effect measure ***for which there is sufficient justification***, the wording should follow this template ..." ***Proposed change 4:*** *Last bullet point (Requirement for JCA reporting)*	Thank you. We have edited the text. Not all suggestions are incorporated.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			*"**An** accurate definition ~~(concept, main source of information, measure, timing, effect measure)~~ of any reported outcome **is required and would include a description of the concept, source of information, the measure of the outcome, timing, and effect measure**"*. Rationale: The last bullet is not clear, hence our proposed rewording.	
Hayley Chapman, PFMD	8	209	The document references 'preferably' which does not provide clear direction on whether the measurement of an outcome of interest would be accepted or not. We suggest removing the word "preferably", and where further description is needed it is also provided.	Thank you.
MTE	8	214	Some outcomes may be fully acceptable as support for the risk/benefit ratio assessment of a certain therapy but are less suitable for the needs of JCA. We assume hereby it is referred to the regulatory phase. We would welcome to have this made explicit. Eg. … support for the risk/benefit assessment as in Regulations (EU) 2017/745 and (EU) 2017/746 but might be less suitable ….	Thank you. This issue is already addressed.
Mihai Rotaru, EFPIA	8	215	Proposed change: "… and biomarkers *(see the definitions in Section 3.~~2~~**3**)."* Rationale: Surrogate outcomes and biomarkers were discussed in Section 3.3, not in 3.2.	Thank you. We edited the text.
EFSPI	8	215	**Current wording**: "This may be the case for surrogate outcomes and biomarkers (see the definitions in Section 3.2)" **Proposed wording**: "This may be the case for surrogate outcomes (see the definitions in Section 3.2)" Rationale: surrogate outcomes may include clinical outcomes or biomarkers and we consider that there is no need to call out.	Thank you. We edited the text
Tanja Podkon	8	215	Original text: (see the definitions in Section 3.2)	Thank you. We edited the text

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
jak, Takeda			Comments: the surrogate outcomes and biomarkers were discussed in Section 3.3, not 3.2 Suggested Change (correcting section number)	
Mihai Rotaru, EFPIA	8	216	Proposed change: *"In general, long-term or final outcomes (i.e., the occurrence of an irreversible event of primary interest such as death) are ~~preferred~~* ***important*** *in HTA.* ***However, longer-term outcomes may not be preferable in practice, if the assessment period extends beyond the primary outcome assessment timepoint with reduced causal attribution and higher uncertainty.****"* Rationale: Preferred suggests a ranking of health outcomes, that is now allowed within the HTA. These outcomes are important, but not the only ones relevant to HTA. The term "irreversible event" is used repeatedly – and is believed to be ambiguous and unhelpful. The event needs to be objective and definite, but irreversible seems to imply something about the event's consequences. Similar statements on long-term or final outcomes are made at other places in the document. It is certainly relevant to state that longer-term outcomes are important; however, in practice longer-term outcomes may often not be preferable, especially if time-periods beyond the primary outcome assessment are considered. Both the validity (given drop-out and other intercurrent events) and the uncertainty of long-term outcomes should always be considered and weighed against the benefit of a longer observation time.	Thank you for your input. We will consider it.
European Huntington Association	8	216	Why would irreversible event be preferred? For example, improving quality of life is often reversible but still a positive outcome.	Quality of life is considered as relevant endpoint.
Prof. Matthias P.	8	221	Original wording: "Both the EUnetHTA collaboration and the European Medicines Agency (EMA) have published detailed guidelines…"	We agree that MS makes the the final decision whether to accept an endpoint.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer nat, Laura Könenkamp, Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzens, M.Sc., Steven Krüger, M.Sc., SKC Beratungsgesellschaft mbH			Comment: Referring to regulatory guidelines is not necessarily a beneficial and appropriate strategy when it comes to HTAs. The EMA Guidelines referred to in this sentence can contain specific points of guidance, which might currently not be accepted in all member states due to differences in methodological requirements. If the HTD aims to fulfill the requirements concerning outcome choice and measurement of the EMA guidelines, there should be certainty that the HTA bodies of the separate MS are also accepting the requirements of the proposed guidelines.	
Prof. Matthias P. Schönermark,	8	222	We would like to point out that the links connected to source [14] and [15] are not redirecting the reader to a valid web page.	Thank you

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
M.D., Ph.D., Ingo Hantke, Dr. rer nat, Laura Könenkamp, Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzens, M.Sc., Steven Krüger, M.Sc., SKC Beratungsgesellschaft mbH				
HTAi PCIG	8	Line 226	The document states that: *Deciding what is a patient-centred outcome for the PICO question for a particular therapy should **ideally** be done in close collaboration with patients and healthcare professionals who either live with the medical condition and/or are knowledgeable about the condition. However, the final decision is up to the individual MS."* **The word 'ideally' must be deleted.**	Thank you. We have edited the text

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Deciding what outcomes are patient-centred cannot be done by proxies. Much research has shown that patients and clinicians do not agree in the assessment of the importance of outcomes. The decision on what is a patient-centred outcome for the PICO must be taken with people who either live with the condition (patients) or care for a person with the condition (care givers). Asking healthcare professionals provides 'clinician-centred care' perspectives. Even though it is acknowledged that in many healthcare systems, clinicians play a key role in collecting PROs, their view cannot be considered as representative of patients' ones.	
Hayley Chapman, PFMD	8	226	The document states "Deciding what is a patient-centred outcome for the PICO question for a particular therapy should ideally be done in close collaboration with patients and healthcare professionals who either live with the medical condition and/or are knowledgeable about the condition. However, the final decision is up to the individual MS. We would propose that the word "ideally" and "/or" are removed as it should be done in close collabration with the patients. Patients would then also include healthcare professionals living with the condition. Healthcare professionals who are knowledgeable about the condition will not be able to provide the same input as someone who is living with the disease.	Thank you. No change.
EFSPI	8	229	"However, the final decision is up to the individual MS." EFSPI would welcome clarity on how the final decision by Member States is made. Specifically, we would like to understand what types of evidence are considered when making the decision (e.g. qualitative research with patients) and proposes adding examples of the key evidence needed.	Already addressed issue.
EORTC	8	Line 229	This is a non patient centric comment. It does not make sense, if it is of relevance for the patients and HCPs, MS should align. It may be economically wise in their best interest as well for instance for de-escalation studies. (see general comments of the EORTC)	Thank you for your input.
Edward	9	252-260/	The use of COS is not standardized to all disease areas. Therefore, in our	Thank you for your comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
s Lifesciences		Section 3.2 Determinant outcomes for specific therapeutic areas	perspective, it is much relevant to refer to existing Clinical Guidelines or to the disease specific core outcome set when available.	No change
Edwards Lifesciences	9	261-265/ Section 3.2 Determinant outcomes for specific therapeutic areas	We believe that the described actions of involving the stakeholders, such as patients, caregivers and healthcare professionals to identify the COS should be done very early in advance and not at the time of the JCA. For the sake of transparency and to have an inclusive approach of the key stakeholders, we believe **the HTDs should be involved across the entire process and participate during:** - **the pre-JCA (PICO definition and the scoping meetings, align on the evidence requirements and on the timelines** required for the submission of the dossier for JCA), - **the JCA** (i.e., dossier submission and review of the first draft JCA, and not only for fact checking of the final JCA report), - **the post-JCA timeframe (the use and uptake of the** JCA to inform **timely decision on reimbursement, funding and use of the technologies)** Edwards Lifesciences believes that any assessment should consider the following points to be successful: - Ensure the **predictability** of the process and its outcomes for the innovative technology developers - Ensure JCA accelerates the access for the patients and is linked and secures **reimbursement/coverage decisions** - We would like to encourage innovative approaches in conducting assessment and generating the evidence and **move to a lifecycle approach** in the evidence generation and the assessment methods	Thank you for your comment. The involvement of the HTDs in the HTA process is out of scope of this guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			- The **adaptability of the assessment** could consider one or more of the following 3 dimensions: o **nature of the technology** (i.e., implantable MD, CDx, MDx, digital technologies….) o **nature of the disease** (i.e., CVD, 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.	
Mihai Rotaru, EFPIA	9	Summary Box after Line 250	Summary box: EFPIA propose that this box reflects: 1) Joint Clinical Assessment should include clinical endpoints, patient-centred endpoints (including final endpoints) and safety 2) It is important that decision-making is based on the collective evidence; excluding endpoints that are important for clinical-patient and regulatory decision-making can result in sub-optimal decisions at a Member State 3) When discussing surrogacy, this should recognise that surrogacy is towards all patient-centred outcomes and not just final outcomes.	We will consider if clarifications are needed for the next version of the draft.
GSK	9	236 / 3.1	4- It's not clear what is meant by outcomes that are "long-term or final". Obviously, mortality is a "final" outcome, but for morbidity and other types of outcomes, it's not readily apparent. It would be helpful to provide a clearer definition of these types of endpoints, perhaps through examples.	Thank you. This issue is already addressed.
Roche	9	Box: Points of attention for the assessm	Current wording: The EUnetHTA guidelines recommend that outcomes relevant for HTA should be long-term or final where possible. Proposed rewording: The EUnetHTA guidelines recommend that outcomes	We do not think this proposition is useful.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
		ent scoping process	relevant for HTA should be long-term or final where possible. **However, the research question, the disease and treatment investigated will be the most important when deciding on the relevance of different outcomes for PICO questions or JCA, and clinical outcome assessments to directly measure treatment benefit may be appropriate depending on the context of use'.**	
Tanja Podkonjak, Takeda	9	241-250/3.1	Current wording: "*If it is not feasible to measure a final outcome , then intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome [14]"* Suggested wording: ***"The choice of endpoint will depend on the target population and main characteristics of a disease (e.g. non life-threatening versus life-threatening disease) as well as on the aim of therapy (e.g. curative versus palliative therapy).*** *If it is not feasible to measure a final outcome or* ***not appropriate for the condition,*** *then intermediate or surrogate outcomes may be acceptable if* ***they are an established surrogate endpoint or a candidate surrogate which has been accepted by regulators and deemed clinically relevant by patients and clinicians****."* Rationale: Takeda notes that despite referencing the EUnetHTA 2015 guideline, *Endpoints used for Relative Effectiveness Assessment: Clinical Endpoints,* the language in the proposed D4.4 Endpoint guideline, such as the highlighted quote, is stronger and not aligned to the guideline it references.[1] The misalignment is in both language around the consideration of the context and condition in selecting an endpoint but also on the feasability of establishing surrogacy. Furthermore, Takeda recommeds the guideline differentiate between established and candidate surrogate endpoints in their	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			application to JCA and their respective validation requirements as these have different data availability. In specific, Takeda would like to highlight the following quotes from the EUnetHTA 2015 Clinical Endpoints guideline which are contradicted by the proposed text: **Section 2.1 (pg 9)** ***The choice of endpoint will depend on the target population and main characteristics of a disease (e.g. non life-threatening versus life-threatening disease) as well as on the aim of therapy (e.g. curative versus palliative therapy).*** *Final endpoints will typically measure mortality or survival, whereas non-final endpoints measure morbidity and function. Depending on the context, final endpoints (e.g. survival in curative therapy of a life-threatening disease) are preferred, whereas non-final endpoints may be more suitable to assess treatment benefit in other situations (e.g. HRQoL in palliative therapy or symptoms in non-life-threatening symptomatic diseases).* **Section 2.1.2 (pg 10)** *It should be acknowledged that the relationship between a surrogate and the endpoint of interest can never be considered as definite. Even if well established for a given health technology, this relationship may be challenged with another health technology that provides the same effect on the surrogate, but with an unexpected effect on the final endpoint.* **Section 2.1.2 (pg 10)** ***In oncology, PFS is an intermediate endpoint that is relevant on its own right.*** *The use of progression free survival has not the same impact in adjuvant and in metastatic disease. In the adjuvant setting, PFS use appears acceptable; in the metastatic setting, data on PFS alone is insufficient and should be coupled with quality of life assessment and survival data, the maturity of which will be considered on the case by case basis.* Reference: • European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Clinical Endpoints. Adapted version	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			2015.	
Mihai Rotaru, EFPIA	9	251-280	EFPIA welcome the consideration of a standardized set of outcomes, defined as "Core Outcomes Set" (COS). The standardised set of outcomes allows for the results across trials to be combined or compared, reduces the potential for reporting bias and ensures that outcomes are meaningful, relevant and useable. Nevertheless, we feel that this section cannot be operationalised currently since COS are, at present, limited and might be relevant in a few complex disease (e.g., rheumatology). The Guidance for developing and reporting high-quality COS is evolving, however a number of methodological uncertainties still remain, and COS may differ even within a disease (e.g. different cancer types or even by stage of cancer). This section would benefit from highlighting some of these issues. EFPIA believes it is important to ensure that the COS is kept continually updated to account for new and emerging outcomes based on alignment with Health Authorities, clinical groups and patients. We believe that the guidance could be clearer in **making an explicit recommendation to call for the development of COS**. This could be included in an additional section "Future Recommendations", added at the end of the draft guideline. We feel that the following recommendation "generic multi-attribute utility instruments should complement the use of COS" is misleading, confusing and could be interpreted as economic health state utility. We therefore propose to delete this sentence unless clarified further. The authors may want to consider the role of a modular approach leveraging item libraries. The opportunity to leverage item libraries to capture core concepts that are relevant to patients is a critical aspect of a COA measurement strategy to ensure fit-for-purpose requirements to clinical study design and data analysis, while also greatly reducing patient burden. Reference: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-020-01197-3	Thank you for your comment. We agree with you as we consider that the sentence in the text "generic multi-attribute utility instruments should complement the use of COS" might be misleading. We will consider your comment in the next version of the guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			The authors may also want to consider adding reference to the FDA draft guidance on "Core Patient-Reported Outcomes in Cancer Clinical Trials": https://www.fda.gov/regulatory-information/search-fda-guidance-documents/core-patient-reported-outcomes-cancer-clinical-trials	
S.Walleser Autiero Medtronic	9	251-265	Consideration of non-drug technologies must be taken into account when applying any form of core outcome sets as a minimum. Some of these outcomes may not be amenable to clinical studies of devices.	Thank you for your comment.
Françoi s Houÿez , Eurordi s	9	236-250	In elderly population, "relative survival" can be used. How acceptable is it for HTA?	We will consider if clarifications are needed for the next version of the draft.
Tanja Podkonjak, Takeda	9	236-250	The current text states that the guideline recommends outcomes relevant for HTA be long-term or final outcomes and then follows this statement with two outcome measures the guideline considers as 'long-term' or 'final': all-cause mortality, overall survival (OS) or mortality rates/survival rates. This section does not consider conditions or treatments which do not impact survival. We request this section and the overall guideline be broadened to consider final or long-term outcomes beyond mortality. Mortality is not applicable to all therapeutic considerations, for an example in treatments of depression. We request the definition be re-considered and broadened to be applicable to a variety of conditions – particularly considering the HTA Regulation will eventually impact all disease areas.	When reading the whole guideline, it seems clear we do not consider mortality as the only relevant outcome.
Roche	9	241-250	Current wording: "If it is not feasible to measure a final outcome , then intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome [14]"	We do not think this proposition is useful.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Proposed rewording: "If it is not feasible to measure a final outcome, then a surrogate may be used, eg a surrogate that has been used for granting marketing authorization by EMA. [14] Rationale: We think that when EMA accepts a surrogate for marketing authorizations it should be considered an established surrogate acceptable for use by HTAb's. When EMA grants full MA on the basis of a surrogate it is no longer possible to generate long term data on the clinical outcome (for ethical and technical reasons).	
Sebastian Werner vfa	9	236-243	*"The EUnetHTA guideline recommends that outcomes relevant for HTA should be long-term or final [14]. All-cause mortality is an outcome that is objective, easy to measure and definite since the final time point is death. Mortality might be measured either as overall survival (OS) or mortality rates/survival rates for a given period (e.g., 1-year mortality or 5-year mortality). For diseases with expected long-term survival, it might be impossible to obtain mature mortality data from clinical trials at the time at which the JCA report is generated. If it is not feasible to measure a final outcome, then intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome [14]"* Long term and final outcomes are surely desirable. However, marketing authorisation of drugs could be based on a positive benefit-risk evaluation using interim results of a study based on dual endpoints one of which demonstrated superiority (e.g. PFS in a cancer study) and the other (overall survival OS) show at least a clear trend. A final OS outcome is rarely available as studies are often ongoing at the time of JCA. Additional endpoints such as PFS2 and time to next anti-cancer treatment could be considered as intermediate endpoints to substantiate the trend observed for an interim OS outcome. Interim analyses may provide a reasonable precise estimate for long term survival (e.g., 5-years survival rate). Statistical guidance is available and established to extrapolate OS accordingly e.g., for cost-effectiveness evaluations. Intermediate endpoints and statistical methods could be useful to decrease uncertainty of long-term outcomes and	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			thus should be considered in the Joint Clinical Assessment. Another important aspect is the maintenance of the study integrity. For instance, when the request for long-term or final outcomes is associated with unplanned unblinding of the study. Such a request would impact the overall type I error rate, distort the final effectiveness outcomes and consequently a full approval potential on which patients count for and provided their informed study consent. The access to medicines should not be a trade-off between regulatory and HTA requirements to the disadvantage of patients The vfa recommends adding the following text to this as follows: *"Decisions about the acceptability of intermediate or surrogate outcomes should strongly consider that final outcomes might not be available at the early time of JCA and that requests for final outcomes can compromise study integrity (e.g., unpanned unblinding) and therefore should not be expected from the HTD (e.g., when studies would have to be unblinded to obtain the final outcome). Statistical approaches to estimate the long-term outcome using modelling approaches should be considered."*	
Matias Olsen, EUCOPE	9	252-258,	The use of a core outcome set is positive on the one hand as it increases predictability and comparability, however reliance on COS could also risk providing a barrier to innovation in the case where they do not sufficiently capture specific benefits (especially for patient subgroups with more specific needs). Therefore, the developer should always have the possibility to provide additional evidence on outcomes that have not been requested as part of the proposed PICO scheme for a given health technology. This is particularly important in case of rare, paediatric, or severe diseases where all potential aspects of value may not be adequately captured by standard outcome parameters. The COS and disease-specific requests should be communicated to the developer at the earliest time possible, ideally at the planning stage of their studies rather than the Scoping stage for JCA, as part of Joint Scientific Consultations. As noted in our comments on deliverable D7.1.1 "practical guideline on the	Thank you for your comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			interaction between HTD and HTAb", it is crucial to allow for dialogues and appropriate interactions between the developer and the HTA bodies to take place in the EU HTA procedure, and the developer should have the opportunity to discuss the proposed outcomes as part of a scoping meeting and contact the assessors with questions when there is a need for clarification.	
Prof. Matthias P. Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer nat, Laura Könenkamp, Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzens, M.Sc., Steven Krüger, M.Sc., SKC	9	239 - 243	Original wording: "For diseases with expected long-term survival, it might be impossible to obtain mature mortality data from clinical trials at the time at which the JCA report is generated. If it is not feasible to measure a final outcome, then intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome [14]." Comment: The consideration of intermediate or surrogate endpoints in the directive for indications with expected long-term survival is generally appreciated. However, there is concern since acceptance of surrogate parameters as early indicators for OS differs among member states and indications. Therefore, an alignment between member states and consistent guidance should be provided, guaranteeing that a final OS outcome can be considered as not feasible in appropriate situations. Also, accepted surrogate outcomes should be aligned between member states (also see comment for section 3.3, regarding "surrogate outcomes" and comment for line 693-695, regarding PFS). Beside indications in which long-term survival is expected, also other scenarios can and should lead to the consideration of a final OS analysis as not applicable or even ethically not reasonable. E.g., an early, strongly visible (statistically supported) survival advantage over another treatment group should be considered as a scenario in which immature survival data, also in combination with appropriate surrogates, can be recognized as relevant outcome data. This enables patients to switch treatment, if the survival advantage gets clear in a very early stage of the study.	Thank you for your comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Beratungsgesellschaft mbH			Furthermore, a surrogate parameter that is already accepted to correlate with a survival outcome in a specific indication or line of treatment, should not only be accepted as clinically relevant when a final OS-analysis is not feasible. In this situation, the clinical relevance of the surrogate is also given without a prerequisite of an unfeasible final OS-analysis.	
Mihai Rotaru, EFPIA	9	243-246	Proposed change: "*Outcome measurements related to patients' response to the therapy can be reported either as morbidity events or in terms of "time to event"* ~~*(in the case of the occurrence of irreversible binary events)*~~ *or as the change in clinical status or symptoms*." Rationale: EFPIA wish to highlight that if "time to event" (TTE) is solely limited to the occurrence of irreversible binary events, this would a priori exclude relevant reversible outcomes such as response, remission or occurrence of symptoms. Such endpoints are frequently analysed using a TTE approach to account for differences in duration of observation between treatment arms. This has been acknowledged in EUnetHTA draft Guideline 4.5 (Applicability of Evidence), whereby it was stated, "A different solution to the problem of different time points is to use a summary effect measure over time, such as repeated-measures analysis of variance for continuous outcomes or Cox regression for time-to-event data in the single studies. The evidence synthesis can then be performed by using the estimates of the summary effect measure (e.g., the hazard ratio)...." As such, EFPIA recommend the deletion of the statement regarding the occurrence of irreversible binary events.	TTE analysis on other outcomes than irreversible binary event present some conceptual issues. We prefer to keep this precision.
Mihai Rotaru, EFPIA	9	241-243	Proposed change: "*If it is not feasible to measure a* ~~*final*~~ ***patient-centred*** *outcome,* ***which can include a final outcome,*** *then* ***an established*** ~~*intermediate or*~~ *surrogate* ~~*outcomes*~~ *may be* ***used*** ~~*acceptable if*~~*.* ***If an established surrogate is not available, then candidate surrogates may be used if*** *there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the patient-centred outcomes* ~~*final outcome*~~ *[14]."* Rationale:	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			We believe it is important to distinguish between ***candidate*** and ***established*** surrogates, acknowledging that the level of certainty about the validity of a candidate surrogate is necessarily lower than for an established surrogate. Candidate surrogates will be particularly important for cases where there is a new endpoint, a new mechanism of action, or first time in a disease area; all of which are likely to occur given oncology, ATMPs and rare diseases will be assessed in the first few years of the Regulation. Furthermore, we request the surrogacy relates to patient-centred outcomes which includes final outcomes as well as effect on morbidity and HRQoL. The use of final outcomes only restricts measures to refelct mortality or survival only, but not other patient-centred endpoints.	
Matias Olsen, EUCOPE	9	241-243	As noted in the guideline, for certain diseases it might be impossible to obtain mature mortality data for clinical trials at the time of which the JCA report is generated. In the absence of long-term outcomes, surrogate outcomes should be accepted when their use is justified by the available evidence. Replace: If it is not feasible to measure a final outcome, then intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome [14].". With: If it is not feasible to measure a final outcome, then intermediate or surrogate outcomes ~~may be acceptable~~ **will be accepted** if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome [14].".	Already addressed issue.
GSK	9	241-243	*"Intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or*	We will consider it for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			intermediate outcome with the effect on the final outcome [14]" In the reference (Endpoints used for Relative Effectiveness Assessment: Clinical Endpoints) and in the related document (Endpoints used for Relative Effectiveness Assessment: Surrogate Endpoints) it is noted that conclusions can be made despite a lack of high surrogacy if the surrogate threshold effect (STE) is considered. This point should also be included here.	
Norbert Gerbsch for IGES Institut GmbH and Health Econ AG	9	241-243	**Comment:** "*If it is not feasible to measure a final outcome, then intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome."* It is welcomed that intermediate or surrogate outcomes are considered acceptable. However, more detail should be provided, with regard to what is considered "a strong assiciation or correlation of effects".	It is not possible within this practical guideline to propose an unique standardized definition of what constitutes an binary appraisal of something that can be assessed by multiple continuous indices in many situations.
Marjorie Morrison, Lymphoma Coalition	9	246-248	**Clinical Evaluation Measurements.** The deliverable points to the application or use of a range of clinical evaluation measurements and scales to capture patient information regarding health status and disease response to therapies. In addition, the document makes note of the need for well-defined information using validated tools for measurement. **We are in agreement that there is a need for the above however, we propose that it would be advantageous to stakeholders (given the mention in the document of outcome measurements as morbidy or changes in clinical status or symptoms in relation to "time to event") if EUnetHTA21 were to consider whether the provision of additional information and clarity specific to the clinical evaluation measurement tools would be of added value.**	We are not sure to fully understand the comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Roche	9	239-241	Current wording: For diseases with expected long-term survival, it might be impossible to obtain mature mortality data from clinical trials at the time at which the JCA report is generated. Proposed rewording**:** For diseases with expected long-term survival, it might not be feasible to obtain mature mortality data from clinical trials at the time at which the JCA report is generated.The expectations around the review of an outcome should take the study design into consideration to avoid inappropriate evaluation of trial data.	We do not think this precision is useful.
Mihai Rotaru, EFPIA	9	239-240	Proposed change: "*For diseases with expected long-term survival,* ***or extensive confounding by subsequent therapies and cross-over****, it might be impossible…"* Rationale: Additional clarification of situations where it may be impossible to obtain long-term survival impact of the intervention.	We will consider if this clarification is needed.
Mihai Rotaru, EFPIA	9	259-260	*"The* ***COMET*** *259 (Core Outcome Measures in Effectiveness Trials) initiative maintains a COS database [22]."* There are also standards for outcomes development and reporting (core outcome sets-STAD and core outcome sets-STAR) that it may be worthwhile mentioning to the suitability of identified core outcome sets.	Thank you for your comment. We will add these two initiatives in the text, as suggested.
EFSPI	9	247-248	**Current wording**: "It is crucial that the "event" is well defined and that only validated tools for measurement are used." Please provide a reference to the guidance around validation of tools to clarify the definition of a validated tool.	There is a whole section dedicated to this issue within the guideline.
GSK	9	239-240	Statement only refers to the issue of OS in cases of long-term survival but does not include the issue of confounding. Suggested rewording **in bold**: 5- "For diseases with expected long-term survival, **or extensive confounding by subsequent therapies and cross-over**, it might be impossible…"	Duplicated comment.
Roche	9	247-248	Current wording: It is crucial that the "event" is well defined and that only	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			validated tools for measurement are used. Comment: Please provide a reference to the guidance around validation of tools to clarify the definition of a validated tool.	
François Houÿez, Eurordis	9	250-251	The points of attention as described are usually correct, however when a surrogate outcome (meaning a fully validated one) is available, why couldn't it be used in the first place, instead of long-term or final ones? For example arterial blood pressure to monitor the efficacy of treatments to lower it is a surrogate marker, and clinical studies measuring the occurrence of stroke or myocardial infarction are no longer needed. Or HIV RNA viral load is now used to measure the efficacy of antiretroviral treatments, AIDS onset or mortality are no longer needed.	These are rare exceptions and we do not think we can tackle this complex issue within the guideline.
M. Ermisch – GKV-SV	9	Sec. 3,2	The entire chapter on COS seems dispensable given that these are associated with various uncertainties, especially with regard to HTA	Thank you for your comment.
Laurent Petit, Leem	9	3.1	Relevant endpoints: Defining relevant endpoints as meaningful endpoints is a key point of the evaluation as it drives acceptability of outcomes as outcomes of interest. Relevant endpoints might vary according to therapeutic areas and disease stages (i.e. in oncology). **The phrasing below suggests that only OS is a relevant endpoint, and any other endpoint would be a surrogate. The relevance of alternative endpoints should not be exclusively conditioned by the evidence of a correlation or association with the "final outcome" (i.e. mortality).** *"The EUnetHTA guideline recommends that outcomes relevant for HTA should be long-term or final. All-cause mortality is an outcome that is objective, easy to measure and definite since the final time point is death. Mortality might be measured either as overall survival (OS) or mortality rates/survival rates for a given period (e.g., 1-year mortality or 5-year mortality). For diseases with*	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			expected long-term survival, it might be impossible to obtain mature mortality data from clinical trials at the time at which the JCA report is generated. If it is not feasible to measure a final outcome, then intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome." **Suggestion:** ***Given the evolving scope and context of clinical developments in oncology, it appears necessary to consider other endpoints than Overall survival (OS) as clinically relevant endpoints, depending on disease stage and therapeutic intent. Especially in early disease stages, OS is sometimes neither appropriate nor feasible as study outcome, and alternative endpoints should be considered as equally appropriate when the clinical relevance of the endpoint in question is supported by the disease context and directly reflects the interest of patients.***	
Laurent Petit, Leem	9	3.1	For some diseases or disease stages, OS would not be the relevant endpoint (ex : early stages in oncology) as the main objective of the clinical trial is to tackle the disease earlier in the therapeutic strategy. In such cases, OS is not relevant as the main outcome and should not be considered as such: DFS or PFS might then be the most suitable relevant outcomes, and should be qualified as relevant endpoints. In some stages of diseases, an equivalence of OS with another relevant endpoint would be suited. Ex: DFS for early stages in oncology **Suggestion:** **It would be appreciated if JCA guidance could characterize the situations in which specific relevant endpoints should be considered, depending on:** - **Therapeutic area (i.e. oncology, orphan diseases)** In rare cancers, difficulty in recruiting patients can lead to single-arm trials or low statistical powering, resulting in challenges demonstrating statistically significant OS benefit	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			- **Stage of disease and intended treatment effect** Increase of clinical developments in early stage disease: as a result mature survival data may be impossible to obtain or compromised by confounding effects in subsequent treatment lines. Confounded treatment effect can also occur in trials where treatment switching from the control to the experimental arm is permitted after progression., and alternative endpoints tailored to this setting may be more appropriate. **Duration of clinical trials** (OS may take more years to be obtained than the length of clinical trial allows ; or if life expectancy of the disease is close to general population. i.e. metastatic prostate cancer) NB: this topic is becoming increasingly critical, as an increasing number of clinical trials are tackling earlier stages of a given diseases (thanks to better testing and diagnostics, prevention policies, curative intent).	
HTAi PCIG	9	General chapter 3.1	It seems that most of this chapter is about outcomes more generally and only little about patient centred outcomes. So, perhaps structure differently to avoid confusion and interpretative variability among the member states: a) general guidance b) patient-centred outcomes c) clinical outcomes	Patient-centred outcome and clinical outcomes are not mutually exclusive. A outcome that is clinically assessed can be relevant for the patient and that is reflected in the general definition of patient-centred outcomes at the beginning of the section. Patient-centred outcomes must not be confounded with patient-reported outcomes.
Bayer	9	3.1	Comment: There are many debilitating chronic indications where OS is no feasible outcome, therefore, the overall prioritization of OS seems not to be appropriate in an indication-agnostic guidance.	When reading the whole guideline, it seems clear we do not mean that OS is the only relevant outcome for all situations.
GSK	9	Section 3.2	As well as COS, this guideline should consider the available clinical guidelines in specific therapy areas. If clinical guidelines accept certain endpoints as being relevant to the therapy area, they should also be considered as suitable endpoints for the joint clinical assessment and by MS.	Thank you for your comment.
Laurent Petit,	9	3.2	**Core Outcome Set :** **There is a need to clarify the purpose, content and weight in evaluation**	Thank you for your comment. We will consider your comment in the next

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Leem			**of the COS.** **COS might also not be available in each disease, and must be up-to-date** in order to be taken into account in evaluation. *"Even though the recommendations from well-established COS should be considered in the selection of outcomes for the assessment scoping process, if such COS are available, it should be noted that COS are not written from a HTA perspective. Therefore, generic multiattribute utility instruments should complement the use of COS."*	version of the guideline. Duplicated comment.
Prof. Matthias P. Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer nat, Laura Könenkamp, Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzen	9	Section 3.2	Original wording: "Efforts are being conducted to identify a standardized set of outcomes that should be measured and reported, as a minimum, in all clinical trials in specific areas of health or healthcare, defined as a core outcome set (COS). Initially, these initiatives were in medical fields such as rheumatology (see the OMERACT initiative) in which disease manifestation is mostly chronic and heterogeneous and affects more than one organ. In these medical settings, defining a set of the most relevant outcomes is highly challenging, which is why there is a need to define COS at an international level. These initiatives have subsequently been applied in various medical fields and healthcare settings" Comment: The definition of indication-specific COS appears reasonable. Nevertheless, the applicability of a core outcome set for rare diseases with often unique features is limited and needs to be specified.	Thank you. This issue has been addressed earlier.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
s, M.Sc., Steven Krüger, M.Sc., SKC Beratungsgesellschaft mbH				
Mihai Rotaru, EFPIA	9	236	*"The EUnetHTA guideline recommends that outcomes relevant for HTA should be long-term or final [14]."* Comment: EFPIA is concerned regarding the clarity and value of long-term outcomes in the current guideline and strong association with all-cause mortality, given the choice of endpoint will be dependent on the population and characteristics of a disease (e.g., an acute episode or treatment in a non-life-threatening disease). The definition of "*outcomes relevant to HTA*" excludes events such as Myocardial Infarction (MI), which are relevant events for HTA independent of their impact on survival. This is then acknowledged later in the document (line 297). It is thus suggested to add a definition of long-term outcomes. In some cases (rare diseases or vaccines), long term outcomes can be documented several years after the 1st market authorization with real-world data. Furthermore, EFPIA strongly recommend all clinically relevant endpoints are relevant to HTA and should be considered in the JCA. The guideline would benefit from further clarification on what is a final outcome; the current brief definition is in line 216, "the occurrence of an irreversible event of primary interest". We therefore recommend the inclusion of further information consistent with the EUnetHTA 2015 guideline, which stated, *"…the choice of endpoint should be justified and relevant for REA purposes. The choice of endpoint*	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			will depend on the target population and main characteristics of a disease (e.g. non-life-threatening versus life-threatening disease) as well as on the aim of therapy (e.g. curative versus palliative therapy). Final endpoints will typically measure mortality or survival, whereas non-final endpoints measure morbidity and function. Depending on the context, final endpoints (e.g. survival in curative therapy of a life-threatening disease) are preferred, whereas non-final endpoints may be more suitable to assess treatment benefit in other situations (e.g. HRQoL in palliative therapy or symptoms in non-life-threatening symptomatic diseases).”	
HTAi PCIG	9	Line 236	‘Long-term’ is relative and should be defined.	We do not think it is possible to come up with one threshold or one standardized definition of what is “long-term”.
Bayer	9	237	Comment: ACM should not be mentioned as the generally appropriate mortality outcome, since it usually is reflective of factors not specific or linked to disease and treatment. Proposal: Please substitute ACM by “disease specific mortality”.	ACM reporting includes disease specific mortality. Thank you for your comment.
EHA	9	242	We suggest replacing “strong association” with “very strong association” to reinforce the concept that the use of surrogate endpoints should be limited to a few exceptions. The same change should be applied in the “Points of attention” section.	We will consider this suggestion for the next version of the draft.
Tanja Podkonjak, Takeda	9	242	Current text: *“If it is not feasible to measure a final outcome , then intermediate or surrogate outcomes **may** be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome.”* Proposed text: *“If it is not feasible to measure a final outcome, then intermediate or surrogate outcomes **are** acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome.”* Rationale: OS results are not available in most cases for oncology medicines at the time	Thank you for your comment. We believe that the original wording is appropriate. Further details on our position on surrogate endpoints can be found in section 3.3.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			of the regulatory filing due to challenges highlighted in Appendix A; impact of cross-over and confounding from subsequent therapies, long follow-up times required. It is imperative that intermediate or surrogate outcomes are included in a JCA and the totality of evidence considered by the JCA assessors, especially particularly if the endpoint is an established surrogate.	
Mihai Rotaru, EFPIA	9	248	Proposed change: *"It is crucial that the "event" is well defined and that only ~~validated~~* ***fit-for-purpose*** *tools for measurement are used. "* Comment: In the context of Clinical Outcome Assessments and particularly PROs, both validity and reliability are important. The authors may want to consider whether the term "fit-for-purpose" rather than "validated" is more relevant, where "fit-for-purpose" has good acceptance in the outcomes field. "Validated" can be ambiguous (which elements of validity? how has it been done?), whereas "fit-for-purpose" is more clearly defined and is being adopted by the FDA in their June 2022 Draft Patient Focused Drug Development Guidance 3 (https://www.fda.gov/media/159500/download).	Clarifications about the relations between validity, reliability, interpretability and fit for purpose will be added in section 5 of the draft.
EFSPI	9	250	**Current wording**: "The EUnetHTA guidelines recommend that outcomes relevant for HTA should be long-term or final where possible." We propose to add "However, the research question and the disease and treatment investigated will be most important when deciding on the relevance of different outcomes for PICO questions or JCA and clinical outcome assessments to directly measure treatment benefit may be appropriate depending on the context of use."	Duplicated comment.
Sebastian Werner vfa	9	250 Box	"*If it is not feasible to measure final outcomes, then intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome."*	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			The vfa recommends adding clarifications on validation criteria to support a "validated surrogate" and a "strong association" or "correlation of effects" on the surrogate or intermediate outcome with the effect on the final outcome. Validation criteria must be feasible and follow a pragmatic approach based on scientific rationale that is harmonized across Member States. They should follow the generally accepted state of scientific knowledge and the international standards of evidence-based medicine. The criteria should consider the medical and technological context ensuring best fit for the data situation. Adapted validation frameworks for chronic and orphan disease as well as ATMP disease areas are necessary to appropriately capture the disease specificities.	
HTAi PCIG	9	Line 251	For the very first time in the document, there is an acknowledgement of the value of patient involvement. "*By involving a wide range of stakeholders, such as patients, caregivers and health care professionals, it is more likely that patient-centred outcomes will be identified."* However, the document restricts the value of this involvement to the development of core outcome sets (a process outside of this HTA process and independent of the HTA agencies). Without meaningful patient involvement in the HTA process when assessing potential outcomes and endpoints, the HTA will be in danger of prioritising outcomes that have little patient relevance and missing the opportunity to identify those outcomes that are most important to patients. To be meaningful, patient input needs to be well-integrated into the scoping and the wider PICO deliberations. Patient communities can have important insights about the population and comparator. For example, it is essential to consult with patients regarding the appropriateness of a comparator to ensure understanding of any limitations such as barriers to accessing the proposed comparator, reasons patients may not use this comparator, and why it might not be a viable alternative among some or many patients. As with all patient input, these insights can offer alternative perspectives to	Thank you for your comment. Inclusions of patients in the whole process of JCAs and JSCs are defined in other guidelines. We consider this issue out of scope.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			clinicians and health departments. **We strongly recommend that patients are meaningfully involved throughout this HTA process from start to finish (JSC, scoping (including endpoint selection), JCA).**	
Roche	9	257	Roche agrees that defining Core Outcome Sets is highly challenging, and suggests adding to line 257 "**High-quality development is critical for COS since the approach taken can lead to variability in the conclusions drawn. Therefore COS should be evaluated and used only if the appropriate quality is demonstrated**." Adding reference to the below sets-STAD and sets-STAR resources would highlight the need to systematically evaluate and assess the suitability of identified COS. Roche notes that since COS typically rely on consensus for the selection of outcome measures, in some instances they are dominated by legacy measures that have not necessarily been developed using current best-practice methods - such as robust qualitative research to support content validity or modern psychometrics. *References:* *The Core Outcome Set Standards for Development (core outcome sets-STAD ref: https://pubmed.ncbi.nlm.nih.gov/29145404/) and* *Core Outcome Set Standards for Reporting (core outcome sets-STAR ref: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002148)*	Thank you for your comment. We will add these two initiatives in the text, as suggested.
EFSPI	9	259	**Current wording**: "The relevance of COS is highlighted when facing prevalent conditions such as cancer and multimorbidity" We suggest adding for clarity that	Thank you for your comment. It is already stated in the text that: "Even though the recommendations from well-established COS should be considered in

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			"There is limited experience to date in the application of COS to clinical effectiveness evaluations by Health Authorities and HTA bodies. Therefore their applicability to the JCA process still needs to be demonstrated."	the selection of outcomes for the assessment scoping process, if such COS are available, it should be noted that COS are not written from a HTA perspective."
Roche	9	260	Roche recognises the COMET initiative to develop robust COS that consider the specifics of HTA and clinical trial science however, COMET is one source. Roche suggests noting in line 260 "The COMET (Core Outcome Measures in Effectiveness Trials) initiative maintains a COS database [22] **as do other sources, for example the International Consortium for Health Outcomes Measurement**" (ICHOM, ref: https://www.ichom.org/patient-centered-outcome-measures/)."	Thank you for your comment. We will add the ICHOM in the text, as suggested.
Tanja Podkonjak, Takeda	9	260	Current text: *"The COMET (Core Outcome Measures in Effectiveness Trials) initiative maintains a COS database [22]."* There are also standards for outcomes development and reporting (core outcome sets-STAD and core outcome sets-STAR) that we recommend adding.	Thank you for your comment. We will add these two initiatives in the text, as suggested.
Hayley Chapman, PFMD	9	262	It is not until page 9 of the document that there is a reference to the value of engaging stakeholders, including patients "By involving a wide range of stakeholders, such as patients, caregivers and health care professionals, it is more likely that patient-centred outcomes will be identified." We note that this refers only to the development of Core Outcome Sets which is a process independent and separate from the HTA. Engaging patients in the actual HTA process is important for not only defining outcomes and impacts, but also to ensure that patients can provide additional context that is important in health-care decision making and HTA deliberations. There is limited reference throughout the document about the valuable role of patients and the patient community in the HTA process which jeopardizes decision-making that reflects the needs of patients.	Thank you for your comment. Inclusions of patients in the whole process of HTA are defined in other guidelines. We consider this out of scope.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			All aspects of the PICO deliberations, including determination of the meaningful patient-relevant outcomes will need sustained and meaningful patient involvement to ensure that the PICO selection reflects the perspectives of people living with the condition.	
Hayley Chapman, PFMD	9	262	The document states "By involving a wide range of stakeholders, such as patients, caregivers and health care professionals, it is more likely that patient-centred outcomes will be identified." At a time when healthcare resources are limited, it is important to align and streamline efforts for healthcare decision-making. With this in mind, it would be helpful to understand how the HTA aligns with relevant regulatory authorities to ensure efforts are not duplicated, and there is a decreased burden on patients. In particular as it relates to data generation, in addition to data processing. This is becoming imperative as the regulators are also looking for insights from the patient community to input into their regulatory decision-making. We refer you to a recent publication (https://link.springer.com/article/10.1007/s43441-022-00432-x) that speaks about the collective value of patient expereince data (PED) and the need for increased stakeholder collaboration in order to streamline the design, generation, collection and use of PED.	Thank you for your comment. Inclusions of patients in the whole process of HTA are defined in other guidelines. We consider this out of scope.
Roche	9	262	Roche recognises the importance of including a wide range of stakeholders and recommends stating HTDs in this process. Editing the current sentence to "By involving a wide range of stakeholders, such as patients, caregivers, health care professionals **and HTD**, it is more likely that patient-centred outcomes will be identified."	Thank you for your comment. We will amend the text as suggested.
Roche	9	264	Roche notes the potential benefit of COS stated by EUnetHTA, regarding reducing heterogeneity in outcome reporting in clinical studies and possibly	Thank you for your comment. We consider that the use of a COS may

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			facilitating meta-analyses. However, COAs should be selected based on their adequacy to assess the concept of interest in the clinical trial context of use and hence identifying and/or implementing an appropriate COS may not be possible. Therefore, Roche suggests deleting the sentence "By contributing to less heterogeneity in outcome reporting in individual clinical studies, COS use may facilitate the conduct of meta-analyses."	facilitate the conduct of meta-analysis, by contributing to less heterogeneity In outcome reporting in the individual studies included in the NMAs.
Mihai Rotaru, EFPIA	9	279	Proposed change: "*Even though the recommendations from well-established COS should be considered in the selection of outcomes for the assessment scoping process, if such COS are available, it should be noted that COS are not written from a HTA perspective. Therefore,* ***other outcome assessments, deemed clinically relevant by patients and clinical experts for the JCA*** ~~*generic multi-attribute utility instruments should*~~ ***can*** *complement* ***or replace*** *the use of COS*." Rationale: The consideration of core outcome sets may indeed be beneficial overall. The possibility to present additional outcomes should remain explicit. COS are often not written from a clinical trial perspective either. Furthermore, utility instruments are intended to be used for economic assessment. As economic assessments are outside of the scope of the Joint Clinical Assessment (JCA), we suggest removing reference to generic multi-attribute utility instruments.	Thank you for your comment. We agree with you about the need for clarification of this wording. We will consider your suggestion for the next version of the guideline.
EFSPI	9	279	EFSPI proposes adding additional text clarifying the limitations of COS in the JCA process, e.g. after "it should be noted that COS are not written from a HTA perspective." We propose to add "When considering the appropriate methodological approach for outcome measure selection within HTA, COS reflect the minimum important concepts that assess the core aspects of a condition and additional measures are often required to holistically measure the patient experience. Hence, COS are not intended as an exhaustive list of outcomes for consideration within HTA evaluations. Therefore, flexibility is expected to be needed in the	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			interpretation and application of COS, to ensure specific aspects of a disease subgroup or therapeutic intervention can still be considered within the JCA, to allow a full evaluation of benefits and risks. Moreover, since COS development does not consider HTA, the COS may not reflect aspects of the condition that are targeted by the therapy being evaluated. It is recommended that key stakeholder groups are specified for COS development, and encourage the consistent involvement of HTDs." Additionally, we recommend explicitly stating "While well developed and valid COS could inform the minimum disease and treatment concepts to capture, COS should not restrict how an outcome is best measured."	
Roche	9	279	Roche suggests adding for clarity that "It should be noted that COS are not written from a HTA perspective and therefore **there is limited experience to date in the application of COS to clinical effectiveness evaluations by Health Authorities and HTA bodies. As such, their suitability to the JCA process still needs to be evaluated. "** Moreover**,** Roche proposes adding additional text clarifying the limitations of COS in the JCA process. Specifically, **"When considering the appropriate methodology for outcome measure selection within HTA, it is highlighted that COS reflect the <u>minimum important</u> concepts that assess the core aspects of a condition. Additional measures can often be required to holistically measure the patient experience for the therapy being appraised, and having a minimum plus additional targeted outcome measures can add burden to patients. Hence, COS are not intended as an exhaustive list of outcomes for consideration within HTA evaluations. Therefore, flexibility is expected to be needed in the interpretation and application of COS, to ensure the JCA can perform a full evaluation. For example, the targeted population or subpopulation, or the therapy's mode of action may not align between the COS and the clinical trial. Moreover, since COS development does**	Thank you for your comment. We will clarify the text in the next version of the guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			not consider HTA, the COS may not reflect aspects of the condition that are targeted by the therapy being evaluated."	
Tanja Podkonjak, Takeda	9	279	Current wording: "*Therefore, generic multiattribute utility instruments should complement the use of COS*." Suggest rewording: "*Therefore, other outcome assessments, deemed clinically relevant by patients and clinical experts for the JCA, ~~generic multiattribute utility instruments~~ can complement or replace the use of COS."* Rationale: Utility instruments are intended to be used for economic assessment. As economic assessments are outside of the scope of the Joint Clinical Assessment (JCA), we suggest removing reference to generic multi-attribute utility instruments.	Duplicated comment.
Edwards Lifesciences	10	281-284/ Section 3.2 Determinant outcomes for specific therapeutic areas	We believe that this guideline should consider all technologies that are within the scope of the HTA Regulation. The core spirit of the HTA Regulation include the: 6- **accelerate the access** for the patients of innovative technologies 7- ensure **equal access** of innovative therapies for all patients across Europe. Therefore, the guideline should **consider all disease areas** and not just some, in spite of the sequence of technologies defined by the HTAR.	Thank you for your input. The guideline is general and should be used for technologies. The methods mentioned are only examples.
Edwards Lifesciences	10	304-308/ Section 3.3 Surrogate outcomes	According to Article 2(6) of the HTA Regulation on the Definition of JCA, "*joint clinical assessment' of a health technology means the scientific compilation and the description of a comparative analysis of the available clinical evidence on a health technology in comparison with one or more other health technologies or existing procedures, in accordance with an assessment scope agreed pursuant to this Regulation, and based on the scientific aspects of the clinical domains of HTA of the description of the health problem addressed by the health technology and the current use of other health technologies addressing that health problem, the description an*d	This issue has been already discussed. Thank you.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			technical characterisation of the health technology, ***the relative clinical effectiveness, and the relative safety*** *of the health technology*;" Therefore, we suggest integrating "**comparative safety**" in this guideline in replacement of "safety" only.	
European Huntington Association	10	Box	Given the first point raised, the sentence 'A validated surrogate outcome should only be used to replace a patient-centred outcome of interest if absolutely necessary:' should be reconsidered as 'A validated surrogate outcome should be used together with a patient-centred outcome of interest'.	Thank you for your comment. We have amended the text in this section to address this point.
Mihai Rotaru, EFPIA	10	292-297	With reference to: *"A biomarker can be defined as a characteristic that is objectively (reliably and accurately) measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to an intervention [27]. Examples include levels of cholesterol and haemoglobin A1c.* *An intermediate outcome is an outcome such as a measure of a function or of a symptom (disease-free survival, angina frequency, exercise tolerance) but is not the ultimate outcome of the disease, such as survival or the rate of irreversible morbid events (stroke, myocardial infarction) [28]."* Proposed wording: We recommend that these are moved earlier in the document when defining different outcomes and referred to in this section. It would be relevant for this section to discuss **established** and **candidate** surrogates (see comments relative to lines 241-250/3.1)	Thank you for your comment. We do not believe that it is appropriate to split surrogates into candidate and established endpoints.
EFSPI	10	292-297	**Current wording**: "A biomarker can be defined as a characteristic that is objectively (reliably and accurately) measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to an intervention [27]. Examples include levels of cholesterol and haemoglobin A1c. An intermediate outcome is an outcome such as a measure of a function or of a symptom (disease-free survival, angina frequency, exercise tolerance) but is not the ultimate outcome of the disease,	Thank you for your comment. We do not believe that it is appropriate to split surrogates into candidate and established endpoints. In addition, we believe that a biomarker is an important concept to define when considering surrogate outcomes.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			such as survival or the rate of irreversible morbid events (stroke, myocardial infarction) [28]." **Proposed wording**: Remove paragraph and define established and candidate surrogates instead. Rationale: We think that these two paragraphs can be removed. Surrogate outcomes may include clinical outcomes or biomarkers	
EFSPI	10	298-303	**Current wording**: " The use of surrogate outcomes in assessment of the clinical added benefit of a health technology can be controversial since the validity of surrogate outcomes has rarely been rigorously fully established [29–32]. Only a few surrogate outcomes have been shown to be true measures of tangible clinical benefit. The guideline "Endpoints used in relative effectiveness assessment: surrogate endpoints" previously developed during EUnetHTA Joint Action 1/2 outlines the methodological issue with the use of surrogate outcomes [14]". **Proposed wording**: "The use of candidate surrogates is controversial since their validity is not yet established" Rationale: We agree that the use of candidate surrogates is controversial since their validity is not yet established. The use of established surrogates should not be controversial. Established surrogates are considered patient-relevant and clinically relevant by EMA, since EMA has granted a full approval based on them. We support the development of a harmonized set of rules that reflect European Member States consensus and alignment with regulatory criteria to validate/qualify surrogates. On this basis, HTA agencies should accept established surrogates. EFSPI would support the development of a list of final outcomes, established and candidate surrogates for each clinical indication [eg FDA table of surrogates]	Thank you for your comment. We do not believe that it is appropriate to split surrogates into candidate and established endpoints.
EORTC	10	Lines 298-303	Surrogacy cannot be dissociated from design: If PFS may be a debatable end-point, using it in intermediate vs delayed access to treatment and looking at OS subsequently may inform not only on the surrogacy but also on the sue of the treatment alongside the evolution of the disease. The approach as described, appears to be rather drug centric than patient centric.	We believe this guideline highlights the importance of patient centred endpoints. However, we have now added additional sentences to clarify that additional surrogate endpoints can also be

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
				requested, e.g. PFS.
Roche	10	298-303	Current wording: " The use of surrogate outcomes in assessment of the clinical added benefit of a health technology can be controversial since the validity of surrogate outcomes has rarely been rigorously fully established [29–32]. Only a few surrogate outcomes have been shown to be true measures of tangible clinical benefit. The guideline "Endpoints used in relative effectiveness assessment: surrogate endpoints" previously developed during EUnetHTA Joint Action 1/2 outlines the methodological issue with the use of surrogate outcomes [14]". Proposed wording: ~~The use of surrogate outcomes in assessment of the clinical added benefit of a health technology can be controversial since the validity of surrogate outcomes has rarely been rigorously fully established [29–32]. Only a few surrogate outcomes have been shown to be true measures of tangible clinical benefit. The guideline "Endpoints used in relative effectiveness assessment: surrogate endpoints" previously developed during EUnetHTA Joint Action 1/2 outlines the methodological issue with the use of surrogate outcomes [14]~~" (deletion of the paragraph) Rationale: The use of **established** surrogates should not be controversial. Established surrogates are considered patient-relevant and clinically relevant by EMA, since EMA has granted a full approval based on them. We support the development of a **harmonized set of rules** that reflect European Member States consensus and alignment with regulatory criteria to validate/qualify surrogates. On this basis, HTA agencies should accept established surrogates.	Thank you for your comment. However, we believe that this is outside the scope of the current guideline.
Liebenhoff, BAH	10	304 - 308	"Safety is a particularly important consideration when using surrogate outcomes. It is important to accurately capture the risk–benefit profile of an intervention. Even if surrogacy has been demonstrated for a specific efficacy outcome, unexpected side effects of that intervention may lead to an increase in mortality or other unfavourable outcomes. Therefore, safety outcomes of interest should be included at the scoping stage. Other considerations regarding safety are addressed in Section 4."	Thank you for your comment. We have deleted this section as this will be covered in section 4.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Safety aspects have to be covered in section "4 Saftey". For that, please delete this part.	
Mihai Rotaru, EFPIA	10	304-308	Proposed change: ~~*"Safety is a particularly important consideration when using surrogate outcomes. It is important to accurately capture the risk–benefit profile of an intervention. Even if surrogacy has been demonstrated for a specific efficacy outcome, unexpected side effects of that intervention may lead to an increase in mortality or other unfavourable outcomes. Therefore, s*~~ ***S****afety outcomes of interest should be included at the scoping stage. Other considerations regarding safety are addressed in Section 4."* Rationale: EFPIA agree that an assessment of the relative safety of an intervention is an important part of the JCA and safety outcomes of interest should be carefully defined during the scoping process. However, the positive risk-benefit profile is assessed by EMA, only in the case of a positive risk-benefit profile marketing authorisation will be granted. EFPIA therefore recommends this part of the paragraph be deleted. "Safety is always an important issue and particularly important when dealing with surrogate outcomes." The sentence implies treatments investigated using surrogate outcomes are essentially less safe, which is not true. It may be that the authors have in mind that many drugs approved based on surrogate outcomes are intended for long duration of use, such as statins, approved based on changes in cholesterol (a surrogate endpoint) are often prescribed for long duration to time to reduce risk of heart attack. Under this scenario, it is particularly important to have a benign safety profile. However, long treatment duration is not always the scenario. However, in other areas, such as cancer, the requirement of a benign safety profile is less applicable.	Thank you for your comment. We have deleted this section as this will be covered in section 4.
EFSPI	10	304-308	The safety assessment should not interfere with the validation of surrogacy. JCA will be undertaken for technologies approved by EMA and therefore with an established positive benefit-risk profile. **Current wording**: "Safety is a particularly important consideration when	Thank you for your comment. This paragraph has been removed as safety is discussed in Section 4.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			using surrogate outcomes. It is important to accurately capture the risk–benefit profile of an intervention. Even if surrogacy has been demonstrated for a specific efficacy outcome, unexpected side effects of that intervention may lead to an increase in mortality or other unfavourable outcomes. Therefore, safety outcomes of interest should be included at the scoping stage. Other considerations regarding safety are addressed in Section 4." **Proposed wording**: Remove paragraph.	
Roche	10	304-308	The safety assessment should not interfere with the validation of surrogacy. JCA will be undertaken for technologies approved by EMA and therefore with an established positive benefit-risk profile.	Thank you for your comment. This paragraph has been removed as safety is discussed in Section 4
Liebenhoff, BAH	10	281 - 284	"Since cancer is the leading cause of death worldwide and the stepwise approach to performing JCA in the HTAR establishes oncological medicines as the first group of therapeutics to undergo JCA, it is important that this document reflects outcomes for assessing the safety and efficacy of new cancer drug therapies." Besides oncological medicines, ATMPs belong to the first group to undergo JCA. ATMP are developed for multiple indications, for that, it is not constructive to focus on oncological medicines in this section.	Thank you for your comment. We will amend it on the next version of the guideline. Note to Antoine: Regarding ATMPs we do not see the necessity to amend the text due to the wide range of products and the specificity of the theme.
Sebastian Werner vfa	10	287-290	*"A surrogate outcome is an outcome that is intended to replace an outcome of interest that cannot be observed in a trial. It is a variable that provides an indirect measurement of effect in situations in which direct measurement of a patient-centred effect is not feasible or practical"* Surrogate outcomes are not only intended to replace a patient-centered outcomes but might be themselves clinically relevant outcomes in regulatory approval or even a patient-relevant endpoint in HTA (cf. Germany), for instance, Relapse free survival (RFS). Thus, these outcomes by itself should be included in the JCA as the HTAR calls for a broad joint clinical assessment of (all) *"health outcomes"* (Article 8 [6]). A list of validated and generally accepted surrogate outcome for EU HTA should be provided alongside with other relevant health outcomes for central disease areas, like this draft guideline's appendix ("specific definitions of outcomes usually used in oncology"). These lists should be regularly updated to ensure keeping up	Thank you for your comment. However, we believe production of this list is outside the scope of the guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			with clinical research progress.	
Tanja Podkonjak, Takeda	10	281-284	Current text: *"Since cancer is the leading cause of death worldwide and the stepwise approach to performing JCA in the HTAR establishes oncological medicines as the first group of therapeutics to undergo JCA, it is important that this document reflects outcomes for assessing the safety and efficacy of new cancer drug therapies."* Comment: Please note that oncology is not the leading cause of death worldwide, it is cardiovascular disease (https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death; https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death) – we recommend changing this statement to cancer being *'one of the leading causes of death'*. Takeda notes the guideline rationale for providing additional guidance for outcomes for assessing safety and efficacy of oncology medicines as they are the first group of therapeutics to undergo a JCA. However, Advanced Therapy Medicinal Products (ATMPs) are also included in the first group to undergo a JCA; we therefore request further guidance be developed on outcomes for ATMPs.	Thank you for your comment. We will amend it on the next version of the guideline.
Mihai Rotaru, EFPIA	10	281-283	Proposed change: *"Since cancer is the **one of the** leading cause**s** of death worldwide and the stepwise approach to performing JCA in the HTAR establishes oncological medicines as the first group of therapeutics to undergo JCA, it is important that this document reflects outcomes for assessing the safety and efficacy of new cancer drug therapies."* Rationale: Cardiovascular disease is the leading cause of death worldwide, with cancer being the second. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death https://ourworldindata.org/causes-of-death#:~:text=Cardiovascular%20diseases%20are%20the%20leading,second%20big	Thank you for your comment. We will amend it on the next version of the guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			gest%20cause%20are%20cancers. Cancer is not the leading cause of death worldwide, it is cardiovascular disease, please revise rationale. If the rationale is based on the sequence of JCA implementation, we request specific guidance for outcomes of assessing safety and efficacy of ATMPs be added to guideline D4.4 Endpoints, as these also will be included in the first phase of JCAs.	
Matias Olsen, EUCOPE	10	295-297	Disease-free survival is normally considered a surrogate endpoint. The distinction between surrogate and intermediate endpoints as suggested in this guideline is unclear and not broadly used.	Thank you for your comment. We have updated the first sentence to clarify that an intermediate endpoint is a surrogate endpoint.
Mihai Rotaru, EFPIA	10	287-288	Proposed change: *"A surrogate outcome is an outcome that is intended to replace an outcome of interest.* ***Often surrogate outcomes are used when it is impractical to study the outcome of interest*** *~~that cannot be observed in a trial~~."* Rationale: In most cases the outcome of interest (e.g. time to death) can still observed, but just not in sufficient numbers to be practical, so a proxy is used (e.g. time to tumour progression).	Thank you for your comment. We believe that the phrase "impractical to study" demonstrates that these still can be observed.
Mihai Rotaru, EFPIA	10	299-300	*"The use of surrogate outcomes in assessment of the clinical added benefit of a health technology can be controversial since the validity of surrogate outcomes has rarely been rigorously fully established [29–32]."* Comment: This statement needs further context. The ability to establish surrogacy is dependent on various factors, including what is the endpoint surrogacy is being established for, is it a candidate or established surrogate, mechanism of action, statistical principles and thresholds being applied, and clinical and biological plausibility. Inability of being able to show statistical surrogacy is not evidence that surrogacy does not exist. In addition, we recommend the end is rephrased to "rigorously established" or "fully established", or "fully established in a rigorous manner".	Thank you for your comment. We have updated the sentence. We have also included an additional sentence clarifying that surrogate endpoints can be requested in addition to patient-centred outcomes which we hope alleviates your concern. However, we believe that it is important to highlight problems with validity of surrogate outcomes. We have provided references in this section which can provide further context.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Moreover, in line with the EFPIA general comment regarding the consideration of the totality of evidence, we believe this statement regarding the use of surrogate outcomes in the assessment of clinical added benefit is inappropriate in the context of the objectives of this guideline. We are concerned that such a statement may guide Member States from requesting certain clinical and health outcomes to enable a JCA. It is suggested to modify or delete the sentence.	
Mihai Rotaru, EFPIA	10	308-309, Box Summary point 1	Proposed change: "Points of attention for the assessment scoping process A **candidate or an established** ~~validated~~ surrogate outcome ~~should only~~ can be used to replace a patient-centred outcome of interest ~~if absolutely necessary~~: *If evidence for a patient-centred outcome is likely to be available, then this should be requested during the scoping process* ***together with surrogate outcomes*** ~~*instead of surrogate outcomes such as morbidity, overall mortality and HRQoL.*~~*"* Rationale: The recommendation should be shortened and moved up to section above on "patient centred" outcomes. (1) in its current form, the statement is confusing as it implies that morbidity, mortality and HRQoL are surrogate outcomes; (2) a "surrogate outcome" (or, more accurately, a clinical outcome) may have informative value of its own for a MS and thus not just "placeholders" for other outcomes. As outcomes assessment may serve different healthcare interests or purposes in different MS, the guidance should not recommend what endpoints (not) to request. (see lines 217-219. "The acceptability of an outcome is subject to MS interpretation of their relevance within their national process for decision-making and thus may differ between MS.") (3) In particular, mortality data (or data on deaths in general) are reported in every study and thus "likely available", but often immature at time of first assessment. In such situation to recommend to not (also) consider "surrogate" endpoints amounts	Thank you for your comment. We do not agree with your proposed sentence. However: (1) We have added an additional comment to clarify that surrogate outcomes can be requested in addition to patient centred outcomes. (2) We have rephrased the first bullet point for clarity. (3) We have included antibody titre after vaccination as an example of a biomarker.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			to recommending MS to refrain from assessing the overall value of a treatment for patients. It should also be noted that surrogates can be used for more than just patient-centered outcomes and be relevant for JCA. For example, antibody titres can be surrogates for disease prevention. Furthermore, surrogate endpoints play an important role in rare diseases and disease areas where no active or disease-modifying treatments exist because there may be no established patient-centred or clinically relevant outcomes.	
Mihai Rotaru, EFPIA	10	308-309, Box Summary point 2	**Points of attention for the assessment scoping process** Proposed change: - delete or reword the following statement: ~~*'Only surrogate outcomes for which validity has previously been clearly established should be requested where possible. This may not be possible at the scoping stage in many instances, although in some cases might have been established by previous JCAs or in other literature on the same indication [14].'*~~ Rationale: This recommendation, as it stands, is almost impossible to be followed; the validation of surrogate outcomes is specific to target disease group and for the intervention under assessment. EFPIA strongly disagree with this contradictory statement and believe that all clinical endpoints that are used for regulatory, HTA and clinical and patient decision-making are relevant for inclusion in JCA (see General comment number 4). Furthermore, some EU HTA bodies require the clinical endpoints from clinical trials, irrespective of surrogacy. Our proposal is that: 1) Clinical endpoints are important outcomes in their own right (as previously outlined) 2) Surrogacy of a clinical endpoint to a patient-centred outcome should be established during the JCA itself	Thank you for your comment. We have included an additional sentence clarifying that surrogate endpoints can be requested in addition to patient-centred outcomes.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			3) The scientific validity of a clinical endpoint should be undertaken during the JCA, acknowledging that different decision-making (including regulators and Member States) have different validity thresholds. 4) No pre-judgement should be made on the surrogacy of an endpoint at scoping	
European Huntington Association	10	3.3	In the current document surrogate outcomes seem to be discouraged or only useable in case any other outcome is not available. Rather than going for an 'if not…/then…' approach, surrogate outcomes should be considered in tangent with PCO during assessment. We understand that rigorous validation has not often been performed. However, promotion of surrogate outcomes will increase the (in)validation of those being considered and can contribute to earlier detection of a disease and/or increased disease prevention.	Thank you for your comment. We have included an additional sentence clarifying that surrogate endpoints can be requested in addition to patient-centred outcomes.
Laurent Petit, Leem	10	3.3	The guidance document focuses on **patient-centred outcomes** (as in the IQWiG doctrine[1]). - Patient-centred outcomes here are assimilated with: HRQoL, morbidity/symptoms, safety, and relying mainly on mortality (line 308 and following). -=> This phrasing is similar to the one in the IQWiG doctrine. There may be a need to move away from such exclusive emphasis centered on mortality, which in some cases will come in contradiction with the previously outlined comments above. "*A validated surrogate outcome should only be used to replace a patient-centred outcome of interest if absolutely necessary:* *If evidence for a patient-centred outcome is likely to be available, then this should be requested during the scoping process instead of surrogate outcomes such as morbidity, overall mortality and HRQoL;* *Only surrogate outcomes for which validity has previously been clearly established should be requested where possible. This may not be possible at the scoping stage in many instances, although in some cases might have been established by previous JCAs or in other literature on the same indication."* **Suggestion :** **The definition of patient-centred outcomes needs to encompass the variations in relevance according to disease areas and disease stages.**	Thank you for your comment. We have added an extra sentence to clarify that surrogate outcomes can be requested in addition to patient centred outcomes.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			7. IQWiG. Allgemeine Methoden, Version 6.1 vom 24.01.2022	
Laurent Petit, Leem	10	3.3	It is recommended that **the guidance document state which outcomes may be considered depending on what circumstances, using examples, such as PFS, DFS, EFS in earlier lines in oncology.** It should be clarified under what conditions **"surrogate" endpoints would become patient-relevant and patient-centred as such, depending on the disease context.**	Thank you for your comment. We believe it is outside the scope of this guideline to provide details on specific disease areas.
Prof. Matthias P. Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer nat, Laura Könenkamp, Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzen	10	Section 3.3	Original wording: "Surrogate outcomes" Comment: Although the consideration of surrogate endpoints in this deliverable is generally appreciated, the uncertainty concerning the acceptance of surrogates in the EU-HTA process is still relatively high. We therefore recommend establishing clear guidelines on what surrogates are considered valid in which scenario. Since according to HTA Regulation (EU) 2021/2282 Article 8 (6) sentence 2 "The assessment scope for joint clinical assessments should be inclusive and should reflect all Member States' needs in terms of data and analyses to be submitted by the health technology developer.", these guidelines should be agreed upon by all MSs.	Thank you for your comment. However, we believe that this suggestion is outside the scope of this guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
s, M.Sc., Steven Krüger, M.Sc., SKC Beratungsgesellschaft mbH				
M. Ermisch – GKV-SV	10	281 ff	It is understandable that the stepwise approach to European joint HTA starting with oncology products and ATMP suggests special advice for this therapeutic area. Consideration should be given to the fact that the appendix might raise the expectation for future provision of additional appendices for other therapeutic areas. While this might be beyond the scope of this guideline, it might be a starting point for specific guidelines reducing the need for individual scientific consultations by answering FAQs. Specific advice on definitions for outcomes typically used in oncology must also contain information on their validity to be helpful for HTD and HTA agencies. Thus, the final sentence of the paragraph should be changed: "Specific definitions of outcomes typically used in oncology **and a synoptical review of their general validity** are provided in Appendix A"	Thank you for your comment. We will consider it for the next draft.
François Houÿez, Eurordis	10	281	Is cancer really the leading cause of death? Cardiovascular diseases, ischaemic heart disease followed by stroke, are the leading cause of death worldwide according to WHO.	Thank you for your comment. We will consider rephrasing the sentence.
Bayer	10	282	Proposal: Please rephrase: "…oncological medicines *are among the first groups* of therapeutics along with ATMPs". Since ATMP are a diverse class of therapeutics for multiple indications of different medical needs, a core outcome set for ATMPs seems to be not feasible.	Thank you for your comment. ATMPs are used in oncology as well in others therapeutic areas, so we believe the wording is correct. .
Mihai	10	Summary	***Proposed change:***	Thank you for your comment. We agree

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Rotaru, EFPIA		Box after line 284	"~~*Generic multiattribute utility instruments should complement the use of COS*~~ ***Other outcome assessments, deemed clinically relevant by patients and clinical experts, can complement or replace COS***" Rationale See above comment. In particular, utility instruments are intended for economic assessment.	with you about the need for clarification of this wording. We will consider your suggestion for the next version of the guideline.
Hayley Chapman, PFMD	10	284	The document states "In the selection of outcomes, recommendations from well-established COS should be considered, if such COS are available." However there is no consideration or recommendation if well-established COS are not available, leaving a lack of clarity and misinterpretation of the recommendations. Also, if they exist, Core Outcome Sets are by their definition a subset of the outcomes that can be measured in a clinical study. There is no consideration given to how additional outcomes will be assessed to check their patient-relevance. So, even if a COS exists for a condition, patient engagement will be needed to further assess and validate additional measures and outcomes captured in the clinical study.	Thank you for your comment. We agree with you about the need for clarification of this wording. We will clarify in the next version of the draft.
MTE	10	285	Surrogate outcomes. More in depth consideration are needed on the ability of use of surrogate outcomes especially in the field of specific medical technologies. Hereby also the consideration of endpoints with combined outcomes and the ability to link (surrogate) outcome when considering best scientific evidence available. (recital 19) followed by re-assessment over time.	Thank you for your comment. We believe the current guideline sufficiently addresses these areas.
Laurent Petit, Leem	10	287	For defining surrogate endpoints, there is a need to include the "patient-centred" aspect in the definition. *"A surrogate outcome is an outcome that is intended to replace an outcome of interest that cannot be observed in a trial"*	Thank you for your comment. The patient-centred aspect of the definition is discussed in the following two sentences.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Suggestion: broaden the surrogate definition itself or its criteria of acceptability, to encompass patient-relevant experiences and not only discuss correlation to relevant endpoints, i.e tumor burden impacting QoL.	
European Huntington Association	10	288	Definition of 'not oberserved'? There is a clear trend towards the wet biomarkers as an outcome as they are more reliable. These can always be observed. Wording should be reconsidered here.	Thank you for your comment. Biomarkers are already referred to in the guideline and we believe that they have been dealt with sufficiently.
European Huntington Association	10	288	Remove 'indirect'. The variable give a direct measure.	Thank you for your comment. We believe direct is most appropriate.
Mihai Rotaru, EFPIA	10	289	It is important that the guideline reflects that surrogate outcomes are important in more settings than just when a patient-centred effect is impractical to study. For example, immunologic response can be a surrogate for prevention of infection or disease, and it is not always possible to directly measure either efficacy or effectiveness. It is now impossible to directly measure efficacy against many vaccine-preventable infections (eg, diphtheria, tetanus, Hib, etc) because there are too few cases in any population as a result of routine vaccine schedules. Therefore, immunologic measures are accepted as surrogates of protection.	Thank you for your comment. We expanded the section on biomarkers to address your comment.
EHA	10	289	We suggest removing "practical", as is it too broad and could leave space for personal interpretation.	Thank you for your comment. Although we agree that practical is a broad description, we believe that it is relevant when describing surrogate outcomes.
Norbert Gerbsch for IGES Institu	10	292 295	**Comment:** *"A **biomarker** can be defined as a characteristic..."* *"An **intermediate outcome** is an outcome such..."* Both are generally considered to be surrogate outcomes. This wording / definition is prone to misunderstandings. We therefore suggest clarifications to avoid misunderstandings if these	Thank you for your suggestion. This has been added.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
t GmbH and Health Econ AG			definitions are used "standalone": **Suggestion:** "A biomarker *is a surrogate which* can be defined as a characteristic..." "An intermediate outcome is *a surrogate* outcome.."	
M. Ermisch – GKV-SV	10	294	The examples chosen for biomarkers are questionable. HbA1c as well as cholesterol are clinical progression parameters used for treatment management. However, proportionality of changes in these to clinical outcomes is not convincingly proven. Thus, either, more convincing biomarkers should be chosen as examples or the examples should be contextualised by changing the sentence to "Examples include levels of cholesterol and haemoglobin A1c, although their surrogacy is disputed".	Thank you for your comment. This section provides examples of biomarkers. We have not provided a conclusion on their validity.
Liebenhoff, BAH	10	295	"An intermediate outcome is an outcome such as a measure of a function or of a symptom (disease- free survival, angina frequency, exercise tolerance) but is not the ultimate outcome of the disease, such as survival or the rate of irreversible morbid events (stroke, myocardial infarction)" In case of non-fatal indication that cause serious symptoms such as pain, without higher mortality it is not constructive to speak from intermediate outcome. For that, please delete this sentence.	Thank you for your comment. This guideline discusses the strengths and limitations of various outcomes and intermediate outcomes are relevant to this discussion.
Bayer	10	295	Comment: The phrase intermediate outcome seems to be misleading as it downgrades serious and patient-relevant morbidity in non-fatal indications that cause serious debilitating symptoms (such as pain) but won´t have a higher mortality (due to pharmacological therapy).	Thank you for your comment. This guideline discusses the strengths and limitations of various outcomes and intermediate outcomes are relevant to this discussion.
Roche	10	298	We propose to add the following text: "**An established surrogate is a surrogate outcome on which EMA has granted full marketing authorization.**"	Thank you for your comment. A broad statement such as this may not apply in all cases for HTA and therefore we do not believe it would be appropriate for this guideline to make such a statement.
Dr. Martin Danner	10	300	After (29-30) should be added: "Surrogate outcomes and biomarker only can be accepted as measures for a patent- centred outcome, if there is a stable evidence for the causal	Thank you for your comment. We believe that this guideline has already provided sufficient guidance on the relationship

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
BAG SELBSTHILFE Germany			connection between the surrogate outcome/biomarker and the clinically / technologically / patient – reportable outcome". Comment: Surrogate Outcomes and Biomarker can't measure patient relevant outcomes	between surrogate outcomes and patient-centred outcomes.
Mihai Rotaru, EFPIA	10	300	**Establishing surrogacy in oncology** We recognize the inherent limitations of the use of surrogate markers in the evaluation of medicines, and particularly in cancer, rare diseases and ATMP. Yet, we understand their importance when few curative therapies are currently available. The limitations of surrogate markers alone should never be used to prevent or slow access to a promising treatment, consistent with what we understand to be the priorities of our cancer advocate partners. In this regard, we commend a useful commentary in the British Medical Journal from Roger Wilson, founder of Sarcoma UK, suggesting the catalytic role of PROs and surrogates: https://blogs.bmj.com/bmj/2021/09/15/surrogate-endpoints-need-complementary-patient-reported-outcomes/. Others in the patient advocacy community have put forward interesting thinking which puts their priorities at the centre when thinking about surrogates in the context of both regulatory and HTA decision making.	Thank you for sharing this information with us.
Mihai Rotaru, EFPIA	10	300	Proposed change: *'Only a few surrogate outcomes* ***(e.g. xxx)*** *have been shown to be true measures of tangible clinical benefit.'* Rationale: To ensure that clarity is provided on how to interpret this sentence, examples would be useful as well as a few references to support such a strong statement. It should also be noted that surrogacy is dependent on a range of issues, for example, tumour type and mechanism of action, and therefore this should not be a blanket statement on any particular endpoint.	Thank you for your comment. It is outside the scope of this guideline to deal with specific examples. New evidence could be published which questions validity of certain outcomes and could cause this guideline to become out of date quickly. In addition, validity of surrogate outcomes should be considered on a case by case basis considering the population, indication etc.
Bayer	10	304	Proposal: The safety endpoints should be covered exclusively in Section 4.	
AIM – Internat	10	308 (frame)	« Only surrogate outcomes for which validity has previously been clearly established should be requested ~~where possible~~. »	Thank you for your comment. We believe that it is important to acknowledge that this

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
ional Association of Mutual Benefit Societies				may not always be possible.
EFSPI	10	308	**Points of attention for the assessment scoping process.** Often so called surrogate endpoints are not intended to replace a patient-centered outcome but are important to supplement or complement the assessment of the patient-centered outcome. They could add value as they could limit gaps related to the certainty of the outcome. For example PFS, PFS2, time to next subsequent anticancer therapy could be intermediate endpoints that could inform about the life-expectancy of the patients and are important for treatment decisions. Thus, please consider to add the following bullet: "Surrogate endpoints could be used to complement the added benefit assessment, in particular when such endpoints are labelled in the prescribing information"	Thank you for your comment. We have added a sentence to clarify that surrogate outcomes can be requested in addition to patient-centred outcomes where relevant.
Laurent Petit, Leem	10	308	*If evidence for a patient-centred outcome is likely to be available, then this should be requested during the scoping process instead of surrogate outcomes such as morbidity, overall mortality and HRQoL;* Or : *If evidence for a patient-centred outcome is likely to be available, then this should be requested during the scoping process instead of surrogate outcomes, such as morbidity, overall mortality and HRQoL;* **Does "such as morbidity, overall mortality and HRQoL" refer to the patient-centred outcomes or surrogate outcomes?** NB : the presence of a comma (here in yellow) or its absence impacts the meaning of the sentence.	Thank you for your comment. We have rephrased this sentence to make it clearer.
Laurent Petit,	10	308	Surrogate endpoints :	Thank you for comment. It should be noted that according to the HTAR, the JCA

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Leem			**«** *If it is not feasible to measure a final outcome, then intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome* *A validated surrogate outcome should only be used to replace a patient-centred outcome of interest if absolutely necessary"* The three levels of validation specified here appear to set a strong standard of statistical evidence in order to maximise endpoint acceptability. However, there is variability in evidence, which may not always be available for all diseases, especially in innovative therapeutic strategies. For example, for a new drug, it may prove impossible to perform any direct or indirect comparison to prove validity of a new endpoint, if it has not been collected before in comparable analogue situations. **Suggestion :** **There is a need to move towards an assessment of endpoint acceptability taking into consideration disease context, rather than a strongly pre-defined table of surrogate acceptability and validity.**	should not contain any ranking of health outcomes. This guideline deals with the methodology for selecting and assessing outcomes rather than the acceptability of specific outcomes.
Sebastian Wernervfa	10 09 11	308 (Box) 250 (Box) 332-333	*"If evidence for a patient-centred outcome is likely to be available, then this should be requested during the scoping process instead of surrogate outcomes such as morbidity, overall mortality and HRQoL"* *"Only surrogate outcomes for which validity has previously been clearly established should be requested where possible."* *"The EUnetHTA guidelines recommend that outcomes relevant for HTA should be long-term or final where possible."* *"If it is not feasible to measure final outcomes, then intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome."* *"A surrogate outcome may lead to greater uncertainty surrounding the benefit*	Thank you for your comment. Article 9 1b of the HTAR specifies that the strengths and limitations of the scientific evidence shall be discussed in a JCA. This guideline provides a basis for assessors to highlight the strengths and limitations of surrogate endpoints. Nevertheless, we have included an additional sentence to clarify that surrogate endpoints can be requested in addition to patient-centred outcomes where relevant. Regarding the inclusion of outcomes accepted in pivotal studies, as stated in

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			of the technology under assessment." The draft guideline implements specific definitions of health outcomes which clearly favour "patient-centred outcomes" and penalize "intermediate and surrogate outcomes" (incl. ranking these health outcomes based on their "uncertainty"), potentially limiting the inclusion and acceptance of relevant health outcomes. However, the HTAR calls for a broad joint clinical assessment of *"health outcomes"* (Article 8 [6]), indicating that *all health outcomes* must be considered in the assessment, while ranking of health outcomes is prohibited (HTAR, Recital 28). Importantly, intermediate, and surrogate outcomes are not only de-prioritized but additionally penalized by elevated requirements they must obey to be accepted. The vfa is very concerned that this EUnetHTA approach will limit the inclusion and acceptance of clinically relevant outcomes of the regulatory approval (often deemed "intermediate and surrogate outcomes"). These limitations implemented by the guideline are not in line with the provisions of the HTAR. The vfa strongly calls for deleting and softening the relevant parts of the guideline that can potentially limit the inclusion and acceptance of clinical outcomes relevant for regulatory approval. *All* health outcomes should be considered in the joint clinical assessment. Further, the vfa recommends that outcomes accepted in pivotal studies as the basis for regulatory approval should be considered in JCA as valid clinical outcome measures. The fundamental principle of evidence-based medicine to use the best available evidence to inform health care decisions, should be applied [1]. [1] Sackett et al. BMJ 1996; 312:71 https://www.bmj.com/content/312/7023/71	the D4.2 *Scoping process* guideline, request for outcomes from MS during the assessment scope should not be data driven. Therefore, outcomes should be only based on MS needs (article 8(6)), and not on what was included in pivotal studies.
Tanja Podkonjak, Takeda	10	Summary box below line 308	Bullet point 2 current wording: *"Only surrogate outcomes for which validity has previously been clearly established should be requested where possible. This may not be possible at the scoping stage in many instances, although in some cases might have been established by previous JCAs or in other literature on the same indication [14]."*	Thank you for your comment. We have included an additional sentence to clarify that surrogate endpoints can be requested in addition to patient-centred outcomes where relevant.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Takeda is concerned about this statement and requests it be revisited or removed from the document. It may not be feasible nor possible to follow – the validation of surrogate outcomes is specific to the disease and intervention in question, following the proposed guidance may also need to be established for the severity of disease or line of treatment. Evidence from the same class of the intervention under assessment may be used but still not often easy to obtain. In other instances, particularly with novel mechanisms of action such as ATMPs, a treatment may be first in class or a first novel treatment in many decades, a previously established surrogacy from neither previous within-class or even within-indication would not be available. We strongly disagree with this contradictory statement and would like to reiterate our position that all clinical endpoints that are used for regulatory, HTA and clinical and patient decision-making are relevant and should be considered in an EU JCA.	
Matias Olsen, EUCOPE	10	292 – 294	Additional examples of biomarkers can be added, i.e. anatomic features such as CNS lesions in multiple sclerosis or atrophic lesions in retina conditions.	Thank you for your comment. We believe that we have already provided a sufficient number of examples.
Edwards Lifesciences	11	323-326/ Section 3.3 Surrogate outcomes	Usually, the association between the surrogate outcomes and the patient-centered outcome is developed in advance by scientific societies and/or clinical research centers and is not driven by HTD. However, the HTD can provide the studies, as available, that are proving the relation between surrogate and final endpoints.	Thank you for your comment. A sentence has been added to indicate that this can also be provided from the scientific literature.
Mihai Rotaru, EFPIA	11	332-333/3.3	Proposed change: "***An established surrogate can be used in lieu of a patient-centred outcome.*** *A* ***candidate*** *surrogate outcome may lead to greater uncertainty surrounding the benefit of the technology under assessment*". Rationale: **Candidate** surrogates, i.e. those endpoints which are still under evaluation and have	Thank you for your comment. We do not believe that it is appropriate to split surrogates into candidate and established endpoints.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			not yet been established as surrogates come with greater uncertainty. ***Established*** surrogate should not lead to uncertainty surrounding the benefit of the technology under assessment, because per definition established surrogates shall substitute for final outcomes.	
Mihai Rotaru, EFPIA	11	334-336 / 3.3	Current wording: "*There are a number of frameworks that may be useful when assessing surrogate outcomes. These include reports by Ciani et al. [31, 33], Grigore et al. [34] and Bujkiewicz et al. [35] and guidelines on preparing a submission to the Australian Pharmaceutical Benefits Advisory Committee [36].*" Comment/rationale: We acknowledge that there are various frameworks that may be useful when assessing surrogate outcomes and would welcome joint guidance from regulators and payers on how to proceed for the validation of a ***candidate*** surrogate. Industry typically faces two situations: 1) EMA grants ***regular*** approval on the basis of an ***established*** surrogate and 2) EMA grants **conditional** approval on the basis of a ***candidate*** surrogate or on the basis of established surrogate endpoints as well. In the first case it is no longer possible to generate long-term data on the clinical outcome (for ethical and technical reasons, e.g. treatment switching) and we would welcome pragmatic guidance on how industry shall navigate this situation. In the second case additional evidence will be generated until conditional approval is transformed into final regulatory approval: this later evidence could be used to support lifecycle HTA, while the initial evidence could be used to inform a first JCA and may inform local decisions to support conditional or temporary reimbursement schemes. Whilst this information describes various references regarding frameworks for assessing surrogate outcomes, this section provides no descriptive information on the use of such frameworks. EFPIA are concerned that such limited information does not provide clarity in the context of the users (e.g., assessors/co-assessors) and objectives of this guideline. Additionally, we suggest reviewing the positioning	We believe that it is outside the scope of this guideline to make such recommendations.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			of this information (currently after the summary box).	
EFSPI	11	333/ Requirements for JCA reporting	**Current wording**: "The assessor should report: ● The level of evidence for the association between the surrogate outcome and the final patient-centred outcome. ● Details on whether this association is based on biological plausibility and/or empirical evidence. ● A description of whether this association has been studied in the disease stage, population and intervention of interest. ● In cases for which the association between the surrogate outcome and the final patient-centred outcome has previously been examined but for a different disease stage, population or intervention, the assessment report should consider the implications for the validity of this association in the current population and intervention of interest. ● The strength of the association between the surrogate outcome and the patient-centred outcome. ● The strength of the association between the treatment effect on the surrogate outcome and the patient-centred outcome. ● Any uncertainties associated with the evidence, and quantified if available. ● The limitations of the use of a surrogate outcome should be explicitly explained. ● Details of any additional information required that could decrease the uncertainty surrounding this outcome. ● An indication of whether or not a patient-centred outcome is likely to be available at a later date. ● Clearly outline any remaining areas of uncertainty. " **Proposed wording**: "The assessor should report: Whether the outcome is considered an established or candidate surrogate along with the rationale". Rationale: We would welcome a clear position on whether the endpoint is considered an established or surrogate endpoint along with the rationale.	Thank you for your comment. We believe that the current list of reporting requirements is sufficient for a JCA.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Roche	11	Box Requirements for JCA reporting	Current wording: "The assessor should report: The level of evidence [...] Proposed wording: "The assessor should report whether the endpoint is considered an established surrogate. For non-established surrogates, the assessor should report: The level of evidence [...] " . Rationale: We suggest the assessor should state whether the endpoint is considered an established surrogate or not, if the surrogate is considered established by regulators then the reporting requirements become redundant.	Thank you for your comment. We believe that the current list of reporting requirements is sufficient for a JCA.
Mihai Rotaru, EFPIA	11	323-326	**Association between the surrogate and patient-centred outcome** Proposed change: *"The HTD should demonstrate the strength of the association between the surrogate outcome and the patient-centred outcome and the treatment effect. ~~This is often done via regression analysis for single studies, or meta-regression in the case of multiple studies.~~ Ideally the association will be demonstrated at both the individual level and the trial level.* ***The HTD could also provide justification for using surrogate endpoint based on existing regulatory/EMA guidance recognizing association between the surrogate outcome and patient-centred outcome."*** Rationale: EFPIA recommend the deletion of this statement, given HTDs will have the responsibility of demonstrating the strength of the association between the surrogate outcome and the patient-centred outcome, which can be done by various statistical methods. Moreover, the EMA guidance recognizing association between surrogate endpoint and patient centred outcome (e.g. Appendix 4 to the guideline on the evaluation of anticancer medicinal products in man – condition specific guidance) are available and considered by Industry when designing clinical trials in specific disease. They are used to justify choice of endpoint as part of the marketing authorisation dossier and should also be considered as part of the joint scientific assessment by HTA.	Thank you for your comment. We have included an additional sentence which includes scientific literature as a method of demonstrating the association. We have also included a sentence at the beginning to this section to clarify that this only needs to be demonstrated if a HTA is providing a surrogate as a replacement to an outcome requested by a member state or if no patient centred outcome is available.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Mihai Rotaru, EFPIA	11	327 - 330	Proposed change: ~~*"For all outcomes requested in the assessment scope, the HTD should provide data, regardless of how immature they are. The presence of surrogate outcome data, regardless of their validity, does not change this requirement. For example, if an intervention is expected to impact OS, data on OS should always be presented, even if the length of follow-up or the number of events is insufficient"*~~ Rationale: ▪ Analysis of immature data introduces uncertainty to trial results, especially for OS. If the number of anticipated events is not reached to perform an analysis, results can be misleading and can lead to wrong conclusions. ▪ The integrity of the study and the data needs to be kept. Publishing immature, uncertain data before the planned read-out can affect the attitude of trial staff and patients to the trial and can cause bias in treatment comparisons. Particular care should be taken to protect the integrity of the trial and to manage and limit appropriately the sharing of information. This concern is highlighted in ICH guidance on statistical principles for clinical trials[1]. There is therefore a clear scientific need to ensure confidentiality of clinical data in these circumstances. ▪ Especially for OS, if a trial reads out positive based on an intermediate co-primary endpoint, the integrity of the data must be retained to able to perform appropriate statistical analysis and draw correct conclusions from the analyses. Immature data may lead to biased results and wrong conclusions. ▪ Data Monitoring Committees (DMC) [1,2] are therefore established to give recommendations to the HTD to disseminate interim analysis results. If the recommendation from the DMC is against disseminating the study results, this recommendation should not be undermined by the request to submit immature data for a publicly available JCA report. References: 1. ICH Topic E9 Statistical Principles for Clinical Trials. September 1998 CPMP/ICH/363/96. 2. CHMP Guideline on data monitoring committees. July 2005 EMEA/CHMP/EWP/5872/03 C	Thank you for your comment. This has been amended to say the latest data cut.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
EFSPI	11	327-330	**Current wording**: "For all outcomes requested in the assessment scope, the HTD should provide data, regardless of how immature they are. The presence of surrogate outcome data, regardless of their validity, does not change this requirement. For example, if an intervention is expected to impact OS, data on OS should always be presented, even if the length of follow-up or the number of events is insufficient." **Proposed wording**: Remove paragraph. Rationale: · Analysis of immature data introduces uncertainty to trial results, especially for OS. If the number of anticipated events is not reached to perform an analysis, results can be misleading and can lead to wrong conclusions. · The integrity of the study and the data needs to be kept. Publishing immature, uncertain data before the planned read-out can affect the attitude of trial staff and patients to the trial and can cause bias in treatment comparisons. Particular care should be taken to protect the integrity of the trial and to manage and limit appropriately the sharing of information. This concern is highlighted in ICH guidance on statistical principles for clinical trials [1]. · Especially for OS, if a trial reads out positive based on an intermediate co-primary endpoint, the integrity of the data must be retained to able to perform appropriate statistical analysis and draw correct conclusions from the analyses. Immature data may lead to biased results and wrong conclusions. References: 1. ICH Topic E9 Statistical Principles for Clinical Trials. September 1998 CPMP/ICH/363/96.	This has been amended to say latest available data cut.
Matias Olsen, EUCOPE	11	331-334	It should be clarified that the association between the surrogate outcome and the patient-centred outcome could be done using the existing literature, in which case no further analysis needs to be done.	Thank you for your comment. We have included an additional sentence which includes scientific literature as a method of demonstrating the association.
EFPIA	11	327-330	Current wording: "For all outcomes requested in the assessment scope, the HTD should provide data, regardless of how immature they are. The presence of surrogate outcome data, regardless of their validity, does not	Thank you for your comment. This has been amended to say latest available data cut.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			change this requirement. For example, if an intervention is expected to impact OS, data on OS should always be presented, even if the length of follow-up or the number of events is insufficient." Proposed removal of the following wording: ~~For all outcomes requested in the assessment scope, the HTD should provide data, regardless of how immature they are. The presence of surrogate outcome data, regardless of their validity, does not change this requirement. For example, if an intervention is expected to impact OS, data on OS should always be presented, even if the length of follow-up or the number of events is insufficient~~ Rationale: · Analysis of immature data introduces uncertainty to trial results, especially for OS. If the number of anticipated events is not reached to perform an analysis, results can be misleading and can lead to wrong conclusions. · The integrity of the study and the data needs to be kept. Publishing immature, uncertain data before the planned read-out can affect the attitude of trial staff and patients to the trial and can cause bias in treatment comparisons. Particular care should be taken to protect the integrity of the trial and to manage and limit appropriately the sharing of information. This concern is highlighted in ICH guidance on statistical principles for clinical trials1. · Especially for OS, if a trial reads out positive based on an intermediate co-primary endpoint, the integrity of the data must be retained to be able to perform appropriate statistical analysis and draw correct conclusions from the analyses. Immature data may lead to biassed results and wrong conclusions.	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			[note: ~~strikethrough~~ denotes proposed deletion] References: 1. ICH Topic E9 Statistical Principles for Clinical Trials. September 1998 CPMP/ICH/363/96.	
EFSPI	11	334-336	**Current wording**: "There are a number of frameworks that may be useful when assessing surrogate outcomes. These include reports by Ciani et al. [31, 33], Grigore et al. [34] and Bujkiewicz et al. [35] and guidelines on preparing a submission to the Australian Pharmaceutical Benefits Advisory Committee [36]." Comment/rationale: We acknowledge that there are various frameworks that may be useful when assessing surrogate outcomes and would welcome joint guidance from regulators and payers on how to proceed for the validation of a candidate surrogate. Industry typically faces two situations:1.) EMA grants regular approval on the basis of an established surrogate and 2.) EMA grants conditional approval on the basis of a candidate surrogate. In the first case it is no longer possible to generate long term data on the clinical outcome (for ethical and technical reasons, e.g. treatment switching) and we would welcome pragmatic guidance on how industry shall navigate this situation. In the second case additional evidence will be generated until conditional approval is transformed into final regulatory approval: this later evidence could be used to support lifecycle HTA, while the initial evidence could be used to inform a first Joint Clinical Assessment and may inform local decisions to support conditional or temporary reimbursement schemes.	We believe that it is outside the scope of this guideline to make such recommendations.
Tanja Podkonjak, Takeda	11	334-336	Current wording: *"There are a number of frameworks that may be useful when assessing surrogate outcomes. These include reports by Ciani et al. [31, 33], Grigore et al. [34] and Bujkiewicz et al. [35] and guidelines on preparing a submission to the Australian Pharmaceutical Benefits Advisory Committee [36]."*	We believe that it is outside the scope of this guideline to make such recommendations

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Comment: We acknowledge that there are various frameworks that may be useful when assessing surrogate outcomes and would welcome joint guidance from regulators and payers on how to proceed for the validation of both an established surrogate and a candidate surrogate. Takeda requests additional pragmatic guidance to HTD, assessors and co-assessors on how to apply the listed frameworks in the context of a JCA be added to the existing guideline.	
EFSPI	11	317-318	**Current wording**: "evidence demonstrating a consistent association between the surrogate outcome and the final patient-centred outcome (from epidemiological or observational studies);" **Proposed wording**: "evidence demonstrating a consistent association between the surrogate outcome and the final patient-centred outcome (from interventional, epidemiological or observational studies);" Rationale: interventional studies may be used in addition to epidemiological and observational studies.	Thank you for your comment. We have updated the guideline to reflect your suggestion.
EFSPI	11	332-333	**Current wording**: "A surrogate outcome may lead to greater uncertainty surrounding the benefit of the technology under assessment." **Proposed wording**: "An established surrogate can be used in lieu of a final outcome. A candidate surrogate outcome may lead to greater uncertainty surrounding the benefit of the technology under assessment." Rationale: Candidate surrogates, i.e., those endpoints which are still under evaluation and have not yet been established as surrogates come with greater uncertainty. Established surrogate should not lead to uncertainty surrounding the benefit of the technology under assessment, because per definition established surrogates shall substitute for final outcomes.	Thank you for your comment. We believe that the original sentence is an accurate reflection as any surrogate may lead to greater uncertainty.
S.Walleser Autiero Medtro	11	327-328	It is noted that HTDs should provide all data, even if immature, on all outcomes requested. It is not clear what is exactly mean with "immature", could this please be clarified better. It should be considered that information from ongoing studies cannot be retrieved before finalisation by a HTD, and	Thank you for your comment. This has been amended to say the latest available data cut.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
nic			there might be commercial and academic in confidence relevant information that can only be shared in confidence and cannot be made public. Processes and safeguards should be in place so HTD can share in-confidence information.	
Roche	11	317-318	Current wording: "evidence demonstrating a consistent association between the surrogate outcome and the final patient-centred outcome (from epidemiological or observational studies);" Proposed wording: "evidence demonstrating a consistent association between the surrogate outcome and the final patient-centred outcome (from interventional, epidemiological or observational studies);" Rationale: interventional studies may be used in addition to epidemiological and observational studies.	Thank you for your comment. We have updated the guideline to reflect your suggestion.
Roche	11	332-333	Current wording: "A surrogate outcome may lead to greater uncertainty surrounding the benefit of the technology under assessment". Proposed wording: "~~A surrogate outcome may lead to greater uncertainty surrounding the benefit of the technology under assessment~~". Rationale: Using an established surrogate does not per se increase the uncertainty, when used in absence of a final outcome.	Thank you for your comment. We believe that the original sentence is an accurate reflection as any surrogate may lead to greater uncertainty.
Laurent Petit, Leem	11	3.3	In line with the previous comment on variability in evidence, which may not always be available for all diseases, especially in innovative and orphan therapeutic strategies, **there is a need for a guidance specific to innovative medicines pioneering clinical developments in new settings.** This is particularly the case for new medicines for which it may prove impossible to perform any direct or indirect comparison to demonstrate validity of a new endpoint, if it has not been collected before for comparable analogues.	Thank you. This is not within the scope of the current guideline, but mat be included in the future.
EHA	11	314	We suggest replacing "comprises a meta-analysis" with "comprises a high-	Thank you.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			quality meta-analysis". This is because there are examples of low-quality meta-analysis on surrogate endpoints, which should not be used to support the use of surrogate endpoints for HTA purposes.	
EFSPI	11	326	**Current wording**" "The HTD should demonstrate the strength of the association between the surrogate outcome…" **Proposed wording**: "The HTD should provide evidence of the strength of the association between the surrogate outcome…" This proposal is meant to clarify that literature references can also be used to support surrogacy argumentation.	Thank you for your suggestion. We have added a sentence to clarify this.
GSK	11	327	"*For all outcomes requested in the assessment scope, the HTD should provide data"* Will there be a limit to the number of outcomes that can be requested? Deliverable D4.2 states that MS should limit their PICO requests to the extent necessary for their national decision-making, but there could potentially be a large number of outcomes requested.	Thank you for your comment. There will be no explicit limit.
Mihai Rotaru, EFPIA	11	Box after line 333 10th bullet	Proposed change: *"~~An indication of whether or not a patient-centred outcome is likely to be available at a later date~~."* Rationale: It is unclear how knowledge that a patient-centred outcome may be available at a later date will help the timely assessment of the relative effectiveness of new medicines that will be undertaken by the nominated assessors and co-assessors. The authors may also want to add a bullet point that reflects that surrogates can be used for more than just patient-centred outcomes. For example, antibody titres can be surrogates for disease prevention.	Thank you for your comment. We have added antibody titres as an example of a biomarker. We believe future availability of outcomes may be useful information for member states.
Mihai Rotaru, EFPIA	12	347-349/4.2	Proposed change: *"During the assessment scoping stage, MS define their required safety outcomes. If specific adverse events are of interest for MS,* ***clear justification must be given on the reason for this data request and an explicit description provided*** ~~*they should*~~	Outcomes (and population, and intervention, and comparators) are based on MS needs (see D4.2 Scoping process), and there is no need for MS to justify their

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			~~*require these explicitly*~~ *(e.g., symptomatic osteonecrosis of the jaw with bisphosphonates)"* Rationale: EFPIA are concerned the current wording risks MS requesting multiple specific adverse events that may ultimately be uninformative or potentially misleading (i.e., of little interest to most of the MS and/or too few events reported to draw any conclusions). Text has therefore been added to provide assessors and co-assessors with more information to aid the PICO consolidation process and ensure the lowest number of safety outcomes possible.	request.
Mihai Rotaru, EFPIA	12	355 / 4.3	Assessment of patient-reported information related to adverse events is not mentioned in this section. Its absence is conspicuous given the topics in the previous sections. The guidance would benefit from mentioning assessments for patient-reported AEs (for example, PRO-CTCAE) and discuss the relationship to the CTCAE (i.e. no reconciliation expected given different data collection modalities).[1] Reference: 1. Overview of the PRO-CTCAE (cancer.gov)	Thank you, this will be added in the next version.
Matias Olsen, EUCOPE	12	368-374	Please clarify how long after the treatment discontinuation adverse events should still be reported.	Reporting for pharmacovigilance is out of scope of this guideline. No change
Prof. Matthias P. Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer	12	341-345	Original wording: "This guideline is not intended to duplicate the definitions already provided for safety terminology [37]." Comment: The clear definition of safety terminology is welcomed and must be clear guidance for all MS to avoid redundancy and mistakable wording.	The reference 37 could be used for that.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
nat, Laura Könenkamp, Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzens, M.Sc., Steven Krüger, M.Sc., SKC Beratungsgesellschaft mbH				
EFSPI	12	342-344 and 371-372	Adverse reactions etc. are commonly part of the Common Technical Documents (CTDs). Avoiding the use of this terminology will be inconsistent with the CTDs. Later in 371-372 it says there are exceptions. More clarification around the use of these safety terms is needed.	There is one exception. We maintain the original wording.
Sebastian Werner vfa	12	346 - 349	*"During the assessment scoping stage, MS define their required safety outcomes. If specific adverse events are of interest for MS, they should require these explicitly (e.g., symptomatic osteonecrosis of the jaw with bisphosphonates)."* The vfa recommends developing a harmonized approach to adverse events that reflects common Member States needs. Specific adverse events that are	This will be considered for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			only relevant to one Member States should be requested at the national level as part of complementary national clinical analyses if these outcomes are numerous in extent. The request of extensive lists of specific adverse events (cf. Germany), which can drive large numbers of analyses in submission dossiers and JCA reports should be avoided and rather be requested at the national level as part of complementary national clinical analyses. A present analysis shows that the number of requested adverse event analyses in German HTA submission dossiers averages 593 (111-1528). [2] [2] Advanced Medical Services (AMS) New requirements for AMNOG-dossiers: Investigation of considered evaluations in the context of the benefit assessment by IQWiG and G-BA, 01. July 2021; download www.vfa.de/report-amnog-dossier-requirements.pdf	
Natacha Bolanos, Lymphoma Coalition	12	361-364	Should be added *"PRO-CTCAE, is the patient-reported outcome measurement system to capture symptomatic adverse events in patients on cancer clinical trials, designed to be used as a companion to the CTCAE".*	Already addressed issue
Prof. Matthias P. Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer nat, Laura Könenkamp,	12	347-349	Original wording: "If specific adverse events are of interest for MS, they should require these explicitly (e.g., symptomatic osteonecrosis of the jaw with bisphosphonates)." Comment: The definition that specific adverse events need to be requested explicitly and upfront by the member state is seen favorable. Nevertheless, manufacturer should still be allowed to provide data for adverse events of special interest if seen relevant by the pharmaceutical company.	This guideline is intended for assessor and co-assessor and is not made to state what is allowed (or not) in the submission dossier.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzens, M.Sc., Steven Krüger, M.Sc., SKC Beratungsgesellschaft mbH				
Roche	12	347-349	Current wording: During the assessment scoping stage, MS define their required safety outcomes. If specific adverse events are of interest for MS, they should require these explicitly (e.g., symptomatic osteonecrosis of the jaw with bisphosphonates) Proposed wording:During the assessment scoping stage, MS define their required safety outcomes. If specific adverse events are of interest for MS, **clear justification must be given on the reason for this data request** ~~they should require these~~ and an ~~explicitly~~ explicit description provided (e.g., symptomatic osteonecrosis of the jaw with bisphosphonates).	Already addressed issue, no justification from MS.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Rationale: Roche are concerned the current wording risks MS requesting multiple specific adverse events that may ultimately be uninformative or potentially misleading (i.e., of little interest to most of the MS and/or too few events reported to draw any conclusions). Text has therefore been added to provide assessors and co-assessors with more information to aid the PICO consolidation process and ensure the lowest number of safety outcomes possible.	
François Houÿez, Eurordis	12	344-346	The term "safety" can be misleading. All health technologies expose patients to some risk. The term "safety profile" or the section on "safety" in patient leaflet is not satisfactory as it can induce the notion that the technology is safe. The term is the preferred by some stakeholders, nevertheless for medicines regulators evaluate the benefit/risk ratio, not the benefit/safety ratio. Other terms could be: dangerousness, or risk.	We prefer to keep the term safety.
Mihai Rotaru, EFPIA	12	368 -369	Proposed change: ~~*"Any suspected unexpected serious adverse reaction (SUSAR) should be reported."*~~ Rationale: The request for SUSARs should be deleted. SUSARs may be medicine specific, and it would be wrong to compare a SUSAR for a new technology with indirect comparators, whilst at the same time not covering their SUSARs. There are explicit reporting requirements to regulatory bodies for SUSARs, and this information will be made available to JCA assessors through section 2.5, 2.7.3 and 2.7.4 from the Common Technical Document submitted to EMA, as requested in D5.1 JCA Template Guideline. Reporting of SUSAR during the JCA could lead to duplicate reporting considering that SUSAR are already reported to regulatory bodies, notably if the RCT is still ongoing at time of JCA. SUSARs should be duly assessed by regulators, HTDs and clinical experts before it becomes relevant for other processes; not to do this could result in misleading conclusions with implications for patient care.	There is a part of duplicate between submitted EMA document and JCA report (for example, results of a phase III trial which fits the PICO request). This potential duplicate is not a justification to not report the results/data in the JCA.
EFSPI	12	368-369	The guideline requires SUSARs (Any suspected unexpected serious adverse reaction) to be reported. A rationale is not given if/how SUSAR reporting	See previous point about SUSAR

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			could inform relative effectiveness assessments. In principle, SUSARs address pharmacovigilance questions and are a reporting obligation to health authorities by HTD. SUSARs will undergo rigid evaluation by HTD to decide whether a specific event a new safety signal. An specific event that had been reported as SUSAR in the beginning of a study need not necessarily to be reported at the end of trial following the safety evaluation. In summary the SUSAR evaluation are reflected in the current drug label. Serious adverse events are reported by the HTD in the HTA dossiers anyway. Please consider to delete the following sentence: ~~Any suspected unexpected serious adverse reaction (SUSAR) should be reported, even if these are (by definition) not requested during the assessment scoping stage.~~	
Sebastian Werner vfa	12	368-369	*"Any suspected unexpected serious adverse reaction (SUSAR) should be reported, even if these are (by definition) not requested during the assessment scoping stage."* A rationale for the request of SUSAR is not given. Further it is not explained how SUSAR reporting could inform relative effectiveness assessments. Basically, SUSARs address pharmacovigilance questions and are a reporting obligation to health authorities by HTD. SUSARs will undergo rigid evaluation by HTD to decide whether a specific event a new safety signal. A specific event that had been reported as SUSAR in the beginning of a study need not necessarily to be reported at the end of trial following the safety evaluation. Thus, the SUSAR evaluation are reflected in the current drug label. Serious adverse events are reported by the HTD in the HTA dossiers, which makes SUSARs unnecessary. The vfa recommends deleting the requirement for report SUSARs.	Already addressed issue
Mihai Rotaru, EFPIA	12	Section 4	**Safety endpoints and risk of misleading results** EFPIA is concerned that the draft proposal could lead to a substantial number of safety analyses that would be substantially disproportionate to the purposes of the JCA. In particular, allowing Member States to request additional safety analyses beyond those outlined, is likely to lead to many more exploratory safety requests,	Already addressed issue

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			particularly from Germany who have a unique need for these exploratory analyses. Recent publications have shown that the safety section of IQWiG submissions averages nearly 600 analyses across an average of 1.2 PICOs (vfa, 2021). Oddens et al. (2022) reported in their case study "The safety-related exploratory analysis of verubecestat led to 206 statistical analyses for treatments and 812 treatment-by-subgroup interaction tests." Recent analysis by EFPIA in a common NSCLC setting, following the Scoping proposal by EUnetHTA 21, has indicated that across the multiple PICOs identified, there would be thousands of requested safety analyses if German requests are incorporated. EFPIA wishes to stress that these analyses would rarely have been pre-specified and would mostly be post-hoc analyses. In addition to the questionable usefulness to most Member States in reviewing these analyses, such extensive analyses pose considerable Type I errors, or false signals (either positive or negative), increasing the risk of findings due to chance and compromising the interpretability of the analyses at a Member State level. As well as potentially contradicting the EMA, whose responsibility it is to establish the overall risk:benefit of a new technology, the publication of such analyses could lead to sub-optimal treatment decisions by those not aware of the issues due to these potentially unreliable analyses, for which the Co-ordination Group would need take accountability for. Therefore, EFPIA proposes that adverse events (AEs) evaluated in the JCA be focused on the very common (≥1/10) and common (≥1/100) as listed in the draft SmPC[1] plus any serious adverse events and any additional clinically relevant AEs as defined by the Regulator and clinical experts. These AEs are proportionate to the needs of JCA and reflect the most important aspects of safety which are relevant to clinical decisions between the clinician and patient when choosing a treatment. EFPIA proposes that no further safety requests are included in the EU JCA and, if more are needed, these are handled as complementary analyses at a Member State level. In addition, EFPIA recommends that the EUnetHTA safety guideline[2] is updated in order to reflect the following points:	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer

			• A clear description of the remit and scope of the future JCA relative safety assessment would be beneficial to avoid overlap or repetition with the regulatory safety assessment. Provide guidance on cross-functional collaboration and sharing of methodologies/best practices on safety assessment between the regulatory and HTA bodies
			• The guideline should consider including recommendations on:
			  • Systematic process and search strategies to identify all relevant safety data
			  • Process or methodology of the comparison of safety data between technology and comparators, in particular where observational or single arms studies are being utilised
			  • Consistency in safety reporting (i.e., who determines the severity of the AE (patient or physician))
			  • Clearer separation between initial and repeat safety assessment
			• Recommend considering all interventions (pharmaceutical, medical devices, and non-drug therapies) in comparative safety assessments and how to evaluate the relative safety of interventions with data from different groups
			• Advocate for guidance on the assessment of safety, with connected e-Health devices alongside medications, and how tools allowing reporting of AE severity by patients should be used
			
			Reference:
			6. European Medicines Agency. How to prepare and review a summary of product characteristics. https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/product-information/how-prepare-review-summary-product-characteristics
			7. European Network for Health Technology Assessment. Endpoints used in Relative Effectiveness Assessment. Safety. Adapted version 2015
			8. vfa, July 2021, http://www.vfa.de/report-amnog-dossier-requirements.pdf
			9. Oddens BJ, Agaku IT, Snyder ES, et al. Exploratory analyses of clinical trial data used for health technology assessments: a retrospective evaluation, BMJ Open 2022;12:e058146. doi: 10.1136/bmjopen-2021-058146	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Tanja Podkonjak, Takeda	12	361 -265	Pelase note that specific adverse events (AEs) may be coded differently for CTCAE or WHO. One example is the reporting of Cytokine Release Syndrome (CRS) which might use several different scales such as ASTCT with reference below "Harmonizing the Definition and Reporting of Cytokine Release Syndrome (CRS) in Immuno-Oncology Clinical Trials". We request the document provide clarification as to which guidance HTD should follow for a JCA submission dossier.	We already state that other terminology could be used, and 'such as' does not mean exhaustivity. The choice of the terminology is not to be done by assessor/co-assessor. No change.
Edwards Lifesciences	12	337-370/ Section 4 Safety	We regret that this guideline does not refer at any point in time to the relative safety or relative effectiveness. According to Article 2(6) of the HTA Regulation on the Definition of JCA, "*joint clinical assessment' of a health technology means the scientific compilation and the description of a comparative analysis of the available clinical evidence on a health technology in comparison with one or more other health technologies or existing procedures, in accordance with an assessment scope agreed pursuant to this Regulation, and based on the scientific aspects of the clinical domains of HTA of the description of the health problem addressed by the health technology and the current use of other health technologies addressing that health problem, the description and technical characterisation of the health technology, **the relative clinical effectiveness, and the relative safety** of the health technology*;" Therefore, we suggest developing the "**comparative safety**" which is the focus of the JCA together with comparative effectiveness, in this guideline in replacement of "safety" only presented in section 4.	This will be considered for the next version of the draft.
MTE	12	338	The section on SAFETY is unclear as which outcomes are considered to evaluated the relative effectiveness and relative safety. i.e which the complementarity is versus analysis available as part of the regulatory frameworks of the pharmaceutical and medical devices regulations. It should also be recognized that for the technologies in scope for medical technologies, the proposed (serious) adverse event might well be context specific ie. dependant on use in national health systems and eg implanters/ users of technology. We welcome	We are sorry but we do not understand this comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			a further clarification in the field of medical technologies.	
Mihai Rotaru, EFPIA	12	Box after Line 345	Proposed change to *"Requirements for JCA reporting"*: ~~*"Use the term "safety", and not "tolerability" or "toxicity". Use the term "adverse event", and not "adverse reaction", "adverse drug reaction", "side effect", "serious incident", "device deficiency", "adverse device effect" or "adverse effect".*~~ Rationale: The guidance recommends avoiding the use of some of the terms listed above, however, these terms have specific meaning in the regulatory context, therefore it is not wise to remove them in preference to only one alternative. We also note that avoided terms in this section are mentioned in the subsequent sections e.g. adverse reaction for a SUSAR.	As explicitly stated, SUSAR is an exception. We maintain the original wording, with the use of safety.
EFSPI	12	347	It should be clarified what is expected in relation to (numerical) reporting of safety endpoints. It is recommended that the guidelines are aligned with the ICH E9 view that "...safety and tolerability implications are best addressed by applying descriptive statistical methods to the data, supplemented by calculation of confidence intervals wherever this aids interpretation."	It will be considered for the next version of the draft.
Roche	12	363	In addition to the standard CTCAE reporting, Roche highlights the Patient-Reported Outcome version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) which is an important component of patient-centred outcome measurement models and should be included in this Guideline to capture the impact of adverse events on patients. Roche suggests adding in text relating to the PRO-CTCAE, covering points such as: The PRO-CTCAE (ref: https://healthcaredelivery.cancer.gov/pro-ctcae/) was developed by the National Cancer Institute (NCI) to systematically capture symptomatic AEs from a patients perspective and complement clinician rated AEs via the CTCAE. The PRO-CTCAE is a validated item library that is used to characterise presence, frequency, severity, and interference with daily function of 78 participant-reportable symptomatic treatment toxicities (Basch et al 2014; Dueck et al 2015). The 78 treatment toxicities in the item library were selected from 790 AEs in the CTCAE (Basch et al 2014). The 78 PRO-	Already addressed issue

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			CTCAE items contain 124 questions rated either dichotomously (presence/absence) or on a 5-point Likert scale (frequency, severity and interference). The standard PRO-CTCAE recall period is 'the past 7 days'. In clinical studies, the PRO-CTCAE is designed to help understand the treatment-related tolerability of all the treatments under evaluation. The PRO-CTCAE is not intended to be used for safety or to be reconciled with CTCAE ratings (Kim et al. 2017) but rather provides complementary tolerability information of the treatment risk-benefit evaluation conducted based on clinical AEs. HTDs are encouraged to pre-specify AEs of relevance from the PRO-CTCAE item library, based on previous preclinical data and regimen-specific information to create a customised measure. Based on this approach, the number of symptoms chosen can vary, but is typically 6-20. Guidance for selecting which items are relevant can be found in the published literature (Trask et al. 2018). Alternatively, in the event that there is no preliminary information on the treatment's potential symptoms, a more general approach can be employed using the published 'cancer core' that includes 12 symptoms/items from the PRO-CTCAE that should be broadly applicable across cancer treatments (Reeve et al., 2014). *References: Basch, E., Reeve, B. B., Mitchell, S. A., Clauser, S. B., Minasian, L. M., Dueck, A. C., ... & Schrag, D. (2014). Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). Journal of the National Cancer Institute, 106(9), dju244.* *Dueck AC, Mendoza TR, Mitchell SA, et al. Validity and Reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). JAMA Oncol. 2015;1(8):1051–1059. doi:10.1001/jamaoncol.2015.2639*	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Trask PC, Dueck AC, Piault E, Campbell A. Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events: Methods for item selection in industry-sponsored oncology clinical trials. Clin Trials. 2018 Dec;15(6):616-623. doi: 10.1177/1740774518799985. Epub 2018 Sep 19. PMID: 30230365. *Reeve, B. B., Mitchell, S. A., Dueck, A. C., Basch, E., Cella, D., Reilly, C. M., ... & Bruner, D. W. (2014). Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. JNCI: Journal of the National Cancer Institute, 106(7), dju129.* *Kim, J., Singh, H., Ayalew, K., Borror, K., Campbell, M., Johnson, L. L., ... & Kluetz, P. G. (2018). Use of PRO Measures to Inform Tolerability in Oncology Trials: Implications for Clinical Review, IND Safety Reporting, and Clinical Site InspectionsUse of PRO Measures to Inform Tolerability in Cancer Trials. Clinical Cancer Research, 24(8), 1780-1784.*	
Roche	12	368	Proposed change: ~~"Any suspected unexpected serious adverse reaction (SUSAR) should be reported."~~ Rationale: The request for SUSARs should be deleted. SUSARs may be medicine specific, and it would be wrong to compare a SUSAR for a new technology with indirect comparators, whilst at the same time not covering their SUSARs. There are explicit reporting requirements to regulatory bodies for SUSARs. They should be duly assessed by regulators & HTDs and clinical experts before it becomes relevant for other processes; not to do this could result in misleading conclusions with implications for patient care.	Already addressed issue
Tanja	12	368	Current text:	Already addressed issue

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Podkonjak, Takeda			*"Any suspected unexpected serious adverse reaction (SUSAR) should be reported."* Suggested wording: ~~*"Any suspected unexpected serious adverse reaction (SUSAR) should be reported."*~~ Rationale: Takeda suggests that the request for SUSARs to be included in the JCA dossier be deleted. SUSARs will already be made available to JCA assessors through section 2.5, 2.7.3 and 2.7.4 from the Common Technical Document submitted to EMA, as requested in D5.1 JCA Template Guideline as there are explicit requirements from regulatory agencies on their reporting. Furthermore, SUSARs are often medicine specific and it may not be feasible nor appropriate to compare a SUSAR for a new technology with a comparator requested through the PICO if it is not included in the intervention clinical trial.	
Tanja Podkonjak, Takeda	13	381-406	Current wording: *"VALIDITY, RELIABILITY AND INTERPRETABILITY OF SCALES"* Suggested rewording: '*VALIDITY, RELIABILITY AND INTERPRETABILITY OF INSTRUMENTS'* and *'5.2 Validity and reliability of instruments'.* Rationale: Instrument is more technically precise. The whole section does not just discuss scales but it broader and covers instruments including scales and items.	To be consistent, we will use the term "outcome measurement instruments"
Mihai Rotaru, EFPIA	13	393-398	Proposed change: *"PROMs (as well as clinically reported measures) can generally be regarded as less objective than performance measures or some technological measures, because they (implicitly or even explicitly) entail subjective appraisal by the patient (or the healthcare professional~~). For example, a performance measure of physical functioning can assess an objective 396 manifestation (e.g., the number of metres a~~*	These considerations will be clarified for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			~~*patient can walk in 6 min), while a PROM item for the same 397 outcome can involve the patient's judgment (e.g., asking the patient if it feels difficult to run 100 m).*~~ ***The selection of COA type should be driven by the concept of interest [1]). Depending on the concept, a PROM may be the most appropriate COA type. For example, in the case of rating pain, this concept is best known to the patient and therefore should be measured via a PRO. For concepts such as physical functioning, multiple COA types may be appropriate to comprehensively measure the concept of interest. For example, the 6 minute walk distance can provide an objective measure of distance walked while a PRO can assess difficulty with ambulation dependent everyday activities, both of which provide important but distinct information****."* Rationale: This statement implies that the performance outcome is the superior measure, when in fact there is value in patient-reported data that is inherently missing from the performance measure. In the example above a patient CAN walk a very good distance in 6min and achieves a satisfactory score for their age category but experiences a great amount of pain when doing so. The "objective" measure is incomplete without the information gained directly from the patient. In addition, that the patient's perception of how far they can walk in 6 minutes is likely to have a more direct relationship with the actual likelihood of walking in their everyday life. In contrast, what they might be able to achieve in a performance test set out in a clinical setting will have poorer association with how much they actually walk on a typical day. EFPIA would like to highlight that whilst PROMs are subjective by their very nature, they are also an important outcome as they measure the direct benefit as perceived by the patient; the instruments selected are fit for this purpose in most circumstances. Recent literature has called for the recognition of PROMs as an objective tool.[2] Reference: *1. Walton MK, Powers JH 3rd, Hobart J, Patrick D, Marquis P, Vamvakas S, Isaac M, Molsen E, Cano S, Burke LB; International Society for Pharmacoeconomics and*	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Outcomes Research Task Force for Clinical Outcomes Assessment. Clinical Outcome Assessments: Conceptual Foundation-Report of the ISPOR Clinical Outcomes Assessment - Emerging Good Practices for Outcomes Research Task Force. Value Health. 2015 Sep;18(6):741-52. doi: 10.1016/j.jval.2015.08.006. Epub 2015 Aug 24. PMID: 26409600; PMCID: PMC4610138. *2. Hamilton, D.F., Giesinger, J. M., and Giesinger, K., Bone Joint Res, 2017, **6**(12) 665-666.*	
Roche	13	393-398	While PROs are more subjective than PerfOs, it is important to reflect that certain concepts are only suitable for self-report by the patient (e.g., pain which is best known to patients directly) and not amenable to other COA measurement approaches. The choice of COA type should be driven by the concept of interest being measured. Proposal to reflect this point by stating **"**PROMs (as well as clinically reported measures) can generally be regarded as less objective than performance measures or some technological measures, because they (implicitly or even explicitly) entail subjective appraisal by the patient (or the healthcare professional). **The selection of COA type should be driven by the concept of interest (Walton et al. 2015). Depending on the concept, a PRO may be the most appropriate COA type. For example in the case of rating pain, this concept is best known to the patient and therefore should be measured via a PRO. For concepts such as physical functioning, multiple COA types may be appropriate to comprehensively measure the concept of interest. For example, the 6 minute walk distance can provide an objective measure of distance walked while a PRO can assess difficulty with ambulation-dependent everyday activities, both of which provide important but distinct information**. Proposal to delete the subsequent example given the above clarification (i.e., 6 minute walk and running 100m) "For example, a performance measure of physical functioning can assess an objective manifestation (e.g., the number of metres a patient can walk in 6 min), while a PROM item for the same	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			outcome can involve the patient's judgement (e.g., asking the patient if it feels difficult to run 100 m)" *Reference: Walton MK, Powers JH 3rd, Hobart J, Patrick D, Marquis P, Vamvakas S, Isaac M, Molsen E, Cano S, Burke LB; International Society for Pharmacoeconomics and Outcomes Research Task Force for Clinical Outcomes Assessment. Clinical Outcome Assessments: Conceptual Foundation-Report of the ISPOR Clinical Outcomes Assessment - Emerging Good Practices for Outcomes Research Task Force. Value Health. 2015 Sep;18(6):741-52. doi: 10.1016/j.jval.2015.08.006. Epub 2015 Aug 24. PMID: 26409600; PMCID: PMC4610138.*	
Mihai Rotaru, EFPIA	13	375-378	Proposed change: *"Causality (attributability) between a heath technology and an AE could be described by many terms and scales. There is* ***always uncertainty and*** ~~*no rationale, and*~~ *a* ~~*high*~~ *risk of bias in* ***the determination of "causality",*** ~~*unblinded studies, to only report AEs potentially related to the health technology under study.*~~ *a safety outcome must always be reported irrespective of causality* ***designation****."* Rationale: The three sentences in this paragraph are true, however the current text does not tie the concepts together in an argument, so putting them in a single paragraph does not make sense. There is also a risk of bias in the causality determination in blinded studies, because blinding is rarely perfect (e.g. difference in AE or clinical response may give the investigator an idea of whether a patient is on control or experimental treatment); the bias is just greater in unblinded studies. Since the conclusion applies to both blinded and unblinded studies, the second sentence does not help.	This will be considered for the next version of the draft.
Matias Olsen, EUCOPE	13	393-396	With respect to yielding results from patient-centred outcomes, especially for areas where the subjective view of the patient is of utmost value (e.g., in case of the impact on feelings or activities of daily life), having a less objective measurement is reflecting the inherent nature of certain types of outcomes. This should be taken into account adequately when addressing the objectivity	Thank you for the comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			of PROMs. Add: "PROMs (as well as clinically reported measures) can generally be regarded as less objective than performance measures or some technological measures, because they (implicitly or even explicitly) entail subjective appraisal by the patient (or the healthcare professional). **However, in some instances (e.g., in case of the impact on feelings or activities of daily life) the subjective appraisal by the patient is the key element of assessing a PROM.**".	
AIM – International Association of Mutual Benefit Societies	13	373-374	« Discontinuation due to an AE (or "adverse event leading to withdrawal") and the type of AE must be reported. Interruption due to an AE and the type of AE must also be reported. »	Thank you. We will consider the input.
Mihai Rotaru, EFPIA	13	373-374	Proposed change: *"Discontinuation due to an AE (or "adverse event leading to withdrawal") **whilst on treatment** must be reported. Interruption due to an AE **whilst on treatment** must also be reported"* Rationale: Clarification of the context of the statement.	Thank you. We will consider it
Mihai Rotaru, EFPIA	13	379-380	EFPIA requests further guidance on exposure adjusted AE analyses in safety reporting for purposes of a JCA. Different lengths of intervention exposure may occur between two arms in a trial, and further guidance on adjustment for exposure length in AE analysis would be useful.	Thank you. The suggestion is out of scope for this guideline.
Mihai Rotaru,	13	390-391	Current wording: "*Outcomes are frequently measured on a continuous scale. The resulting measure*	"Scale" here refers to the ultimate result of the outcome. Items are indeed frequently

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
EFPIA			*can be called a score [43]. Categorical scales are also used."* The first sentence implies that most scores are obtained from outcomes measured on a continuous scale. Most PRO scales have ordinal response options, and through scoring algorithms, ordinal or continuous values are computed, which are generally described as scores in either case.	proposed with ordinal response formats but they are intermediate in the process of producing the measure of the outcome.
EFSPI	13	373-374	**Current wording**: "Discontinuation due to an AE (or "adverse event leading to withdrawal") must be reported. Interruption 373 due to an AE must also be reported." Please precise whether drug discontinuation, trial discontinuation, or both, is meant.	Thank you. We will consider the need for more precise sentence.
GSK	13	373-374	Please clarify that 'discontinuation due to an AE' and 'interruption due to an AE' are in the context of treatment, e.g. 'discontinuation of treatment'.	Duplicated comment. See above
GSK	13	393-394	*"PROMs (as well as clinically reported measures) can generally be regarded as less objective than performance measures or some technological measures"* We respectfully consider this statement to be misleading and may lead to lower value being attributed to PROMs. There has been recent literature calling for the recognition of PROMs as an objective tool (Hamilton, D.F., Giesinger, J. M., and Giesinger, K., *Bone Joint Res*, 2017, **6**(12) 665-666). Moreover, the degree of error in clinicians' measurements compared to PROMs has been found to be larger in some studies (see, e.g., Hahn EA et al., *Mayo Clin Proc*., 2007, **82**(10), 1244-54 where test-retest correlations for heart rate (r=0.68) or blood pressure (r=0.63), compared to physical functioning scale of SF-36 (r=0.93)).	"Objective" and "subjective" are not used here in their colloquial meanings ("true" or "valid" as opposed to "invalid") but refer to the necessity or not of involving a subjective judgment of a person in the process of producing the measure. The purpose of the paragraph was not to discuss measurement properties, nor to discard PROMs, but to emphasize first that there exist measures that do not involve such subjective assessment (e.g., the objective height of a person as measured in cms) vs ones that do (e.g., does a person feel he/she is small or tall?) and secondly that if one wants to have insights on patients' preferences, values, judgments, they need to use a measure which allows the access to these internal standards. Clarifications will be made in the next

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
				version of the draft.
Mihai Rotaru, EFPIA	13	Summary box after line 380	***Proposed change 2*** We suggest adding: ***"Dose reduction due to an AE"*** Rationale: This is an important measure and should be added to the section as well as in the box listing requirements for JCA	Thank you. We will consider it
Tanja Podkonjak, Takeda	13	Summary box below line 380	*"Requirements for JCA reporting"* Comment: Takeda requests further guidance be included in the guideline on the JCA reporting requirements for exposure adjusted AE analyses in safety reporting. Different lengths of exposure may occur between arms in a trial, and further guidance on adjustment for exposure length in AE analysis expected for a JCA would be useful.	Duplicated comment
Mihai Rotaru, EFPIA	13	381	**Validity, reliability and interpretability of the results** There is no mention of the possibility to have a qualification consultation with HTA when new tools are used to measure an outcome (e.g., digital tool) such as the qualification procedure at EMA. This should be envisaged in this guidance in order to consider situation where novel tools or instrument are emerging.	Interactions between HTDs and HTAbs are out of the scope of this guideline.
Roche	13	386	Roche suggests adding a definition of a conceptual framework. The conceptual framework reflects the disease conceptual model - which is a visual depiction of the disease experience (Donatti et al (2008) - plus the measurement model. Therefore the conceptual framework is important to describe in this section. "A conceptual framework outlines the interrelationships among the items and associated domains being measured by the instrument (Rothman et al.	This will be added in the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			2007)." *Reference: Donatti C,Wild D, Hareendran, A(2008) The use of conceptual models, conceptual frameworks, and endpoint models to support label claims of treatment benefit using patient reported outcomes. ISPOR Connect 14, 9–12*) *Rothman, M. L., Beltran, P., Cappelleri, J. C., Lipscomb, J., Teschendorf, B., & Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. (2007). Patient-reported outcomes: conceptual issues. Value in Health, 10, S66-S75.*	
Mihai Rotaru, EFPIA	13	393	Proposed change: *"PROMs (as well as ~~clinically~~ **clinician**- reported **outcome** measures) can…"* Rationale: Suggest using: "clinician reported outcomes measures" instead of "clinically reported measures" in order to be consistent with the term used on Page 6.	This will be clarified in the next version of the draft.
EFSPI	13	393	**Current wording**: "PROMs (as well as clinically reported measures) can generally be regarded as less objective than performance measures or some technological measures, because they (implicitly or even explicitly) entail subjective appraisal by the patient (or the healthcare professional)." It is problematic to suggest that the patient perspective is inherently less objective (and hence, implicitly, less valid) because PROMs often assesses a patient-relevant concept that cannot be addressed by other means. **Proposed wording**: "If a concept can be addressed by both performance measures/technological measures or PROMs (as well as clinically reported measures), the latter may be regarded as less objective because they (implicitly or even explicitly) entail subjective appraisal by the patient (or the healthcare professional)."	These considerations will be clarified for the next version of the draft.
Mihai Rotaru,	13	381 & 406	Proposed change: **"***5. VALIDITY, RELIABILITY AND INTERPRETABILITY OF **INSTRUMENTS** ~~SCALES~~**"***	We will use the term outcome measurement instruments to be more

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
EFPIA			*"5.2 Validity and reliability of ~~scales~~ **instruments**."* Rationale: Instrument is more a technically precise term than scales. The whole section does not just discuss scales but is broader and covers instruments including scales and items.	consistent with the purpose of the guideline.
AIM – International Association of Mutual Benefit Societies	14		Specify~~ing~~ the main source of information as it can have relevance for a given outcome.	This will be corrected.
Mihai Rotaru, EFPIA	14	415-422	**Content validity is missing** This section would benefit from a brief mention of the types of validity (content, construct, responsiveness to change). Internal consistency is not mentioned but may be covered by the mention of structural validity.	References are given for those who wants to have details on measurement properties of outcomes measurement instruments. The sentence does not imply all measurement properties are cited. The guideline is not meant to be a clinimetric or psychometric textbook.
Bayer	6, 8	139-143, summary box	Statement: It is stated that effect measures are used to compare "the measure of outcomes between two intervention groups", but there later is a reference to the "absolute effect". The term "effect" implies a comparison between groups. By definition, any effect is relative as opposed to absolute.	This will be clarified for the next version of the draft.
Mihai Rotaru, EFPIA	14	411-414	Current wording: *"This guideline will also not cover considerations about the measurement properties of routine clinical examination procedures, routine biological and laboratory tests (e.g., measurement of serum creatinine levels), or routine use of medical imaging (e.g., measurement of the size of a particular anatomical structure)."* Comment:	The perimeter of the measurement instruments considered in this guideline will be clarified.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			It is important that the guideline foresees the variety of therapy areas and technologies that could be assessed in the JCA. Given the statement that this guideline will not cover biologic and laboratory tests, we conclude that immune response assays (e.g., antibody titres or functional immune responses) and diagnostic tests (e.g., culture or PCR used to diagnose an infection) are out of scope. If so, then none of Section 5 is applicable to routine vaccines, which may also explain why there is no mention of immunologic responses in other sections of this guideline. Nevertheless, as explained previously, immunogenicity trials are very important to assess vaccines.	
Mihai Rotaru, EFPIA	14	440-443	Current wording: *"Therefore, high inter-rater reliability is paramount (i.e., when the assessment on the same patient is performed by different well-trained professionals, it leads to the same result) [42]. PROMs are completed by patients, so high test–retest reliability has more value for these (i.e., if the assessment is performed by the same patient at two time points with identical conditions, the result is the same) [47]."* Comment: It is not the case that high test-retest reliability has more value than inter-rater reliability for PROs; it is the case that inter-rater reliability is not applicable for PROs at all. Suggest rephrasing the second sentence because it suggests that one type of reliability holds more value than another. The section implies that test-retest reliability is not relevant for ClinROs. There's no reason why test-retest reliability should not be documented for ClinROs as well.	These considerations will be clarified in the next version of the draft.
Ioanna Psalti EUEYE	**14**	411-414	Measures of visual function as adequate primary endpoints when evaluating the safety and efficacy of new ophthalmic drugs include visual acuity, visual fields, colour vision, and contrast sensitivity.[Csaky KG, Richman EA, Ferris FL, 3rd. Report from the NEI/FDA Ophthalmic Clinical Trial Design and Endpoints Symposium. Invest Ophthalmol Vis Sci.	The sentence was misleading. Such outcomes can of course be valid. The perimeter of section 5 will be clarified for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			2008;49(2):479-489]. However visual function assessments may not necessarily pick up biologic activity which anatomic endpoints can measure. It is not clear whether the document accepts such endpoints as lines 411-414 state that the guideline will not cover considerations about the routine use of medical imaging (e.g., measurement of the size of a particular anatomical structure). • Anatomic measures, such as structural endpoints, can be used as surrogate outcomes provided that the validated surrogate implies a result on the true endpoint of interest. Such measures are useful because they provide an objective measurement that can be assessed in the clinic by a number of noninvasive imaging modalities that have been developed specifically for diseases of the eye. Determining the clinical relevance of anatomic endpoints is especially important when testing new therapies for slowly progressing diseases, in which tissue damage can precede vision loss by several years. For example while AMD can affect the fovea from early on, in geographic atrophy its very common for the fovea to be spared until the very last. As a result the conventional endpoint visual acuity (VA), is insensitive to the early stages and slow progression of dry AMD. VA only measures fine visual resolution and underestimates disease progression while scotomas surrounding the fovea enlarge and interfere with reading and other tasks. Effective drug discovery for dry AMD treatments therefore require novel applications of more effective visual function endpoints.	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Mihai Rotaru, EFPIA	14	432-433	Current wording: *"A consensus taxonomy of the psychometric properties of PROMs has been developed by the international Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) group [47]."* Comment: The authors may want to substitute "psychometric properties" with "measurement properties"; content validity (which is listed in the COSMIN list) cannot be established by psychometric work, and further, the title of the cited paper states measurement properties.	This will be done.
Roche	14	408-409	Proposal to clarify that establishing the validity and reliability of scales is relevant to all COA types: "However, in the context of this document, **clinical outcome assessments (i.e., PROs, ObsROs, ClinROs and PerfOs, defined in section 2.1)** are considered"	This will be clarified for the next version of the draft.
Roche	14	425-426	Proposal to modify the text to reflect that for either novel COA development, selection of an existing COA or modification of an existing COA, including qualitative research is fundamental to ensure the adequate assessment of a concept of interest: "**Scale development, modification or use of an existing COA may involve a review of existing patient-centred literature and if needed, qualitative research (e.g, interviews or surveys) with patients, caregivers (where appropriate) and healthcare professionals to identify the most relevant concepts in a given disease or treatment paradigm, and to ensure the selected scale items and instructions are clear, comprehensive and consistently understood."**	We will consider if these clarifications are necessary for the next version of the draft.
Bayer	14	425-426	Proposal: Recommend adding two citations to useful guidance on conducting qualitative studies for PROM development: (i) Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity - establishing and reporting the evidence in newly-developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force Report: part 1 - eliciting concepts for a new PRO	The main purpose of the guideline is not to discuss in depth how outcome measurement instruments must be developed.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			instrument. Value Health. 2011;14(8):967-977. and (ii) Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity - establishing and reporting the evidence in newly-developed patient-reported outcomes (PRO) Instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force Report: part 2 – assessing respondent understanding. Value Health. 2011;14(8); 978-988.	
Mihai Rotaru, EFPIA	14	407	Proposed change: *"For appropriate usage, any measurement ~~device~~* ***instrument*** *needs to meet a sufficient level for two main properties: validity and reliability [42]."* Rationale: The use of the term 'device' is confusing, we recommend that it is replaced with the term 'instrument'.	It will be corrected.
Mihai Rotaru, EFPIA	14	416	Proposed change: *"If a PROM is designed to measure anxiety levels, it must* ***correlate with general anxiety symptoms across conditions but not correlate with other mental health outcomes like*** *~~not measure~~ depression ~~levels.~~"* Rationale: To make the expression more accurate and meaningful to address the underlying ideas of concern. Clarification requested: This statement also excludes multi-domain PROMs with a well-defined subdomain that measures the aspect of interest. Is use of instruments measuring two or more concepts discouraged? If not, we propose that the language is softened to reflect this.	It will be corrected.
Tanja Podkon	14	416	Current wording: *"If a PROM is designed to measure anxiety levels, it must not measure*	Thank you. We will reword the sentence.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
jak, Takeda			*depression levels"* Suggested wording: *"If a PROM is designed to measure anxiety levels, it must correlate with general anxiety symptoms across conditions but not correlate with other mental health outcomes like depression"* Rationale: To make the expression more accurate and meaningful to address the underlying ideas of concern. Clarification Requested: This statement also excludes multi-domain PROMs with a well-defined subdomain that measures the aspect of interest. Is use of instruments measuring two or more concepts discouraged? If not, please soften language.	
Mihai Rotaru, EFPIA	14	422	Proposed change: "*Thus, reliability assesses the extent to which a measure is free from measurement errors (~~i.e.,~~ **e.g.,** random errors)."* Rationale: There can be systematic error as well as random error, so the authors could drop (i.e. random error) at the end of the sentence or mention it as an example.	The sentence is about reliability exclusively. Lack of reliability leads to measurement error while lack of validity can lead to systematic errors.
EFSPI	14	429	**Current wording**: "not one-dimensional properties" **Proposed wording**: "multi-faceted attributes"	It will be corrected.
Roche	14	429	Proposal to change wording from "are not one-dimensional properties" to **"are multifaceted attributes"**	Duplicated comment.
Mihai Rotaru, EFPIA	14	437	Next to line 189, this is the second time the SF-36 is referenced as the sole example. The authors may want to use different examples as the SF-36 is becoming less widely used due to its general nature, when there are more specific and precise measures available.	The document is not about endorsing a list of validated PROMs. Clarifications about generic and specific HRQoL instruments will be added in section 2.

EUnetHTA 21 Public Consultation Comments and Responses
Of D4.4 - Endpoints

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Bayer	14	444	Comment: The term "identical conditions" is quite strict and unattainable in practice. Proposal: Recommend changing wording to *"*PROMs are completed by patients, so high test–retest reliability has more value for these (i.e., if the assessment is performed by the same patient at two time points in the absence of change, the result is the same)*".* This wording is also more consistent with COSMIN terminology.	This will be considered for the next version of the draft.
Mihai Rotaru, EFPIA	14	403–405	*"Distinguishing these differences in perspective in detail and thus the actual outcome collected can require full access to the verbatim items and sometimes even literature on scale development and validation."* Comment: Please note that it may not always be possible for the HTD to provide access to the full verbatim instrument and/or instructions as these may require commercial licenses for use and distribution that forbid such sharing of documents.	Indeed. But this concern is out of the scope of the guideline.
AIM – International Association of Mutual Benefit Societies	15	Frame	Add as a last bullet point of the summary that an instrument's level of validity and reliability does not ensure that a measure of treatment effectiveness has high certainty of results. The result also needs to be put in the context of the study design, conduct and analyses (cf. last paragraph of the section). Add as last bullet point of Requirements for JCA reporting that the subgroup during JCA should also pay attention to the study design while assessing instruments' validity and reliability.	This will be considered if it's necessary for the next version of the draft.
Tanja Podkonjak, Takeda	15	Summary	Current wording: *"References, as provided by the HTD, allowing the access of the specific clinical studies assessing the measurement properties (and measurement model) of the instruments that are used."* Suggested wording: *"References, as provided by the HTD, allowing the access of the specific ~~clinical~~ studies assessing the measurement properties (and measurement model) of the instruments that are used."*	This will be corrected.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Rationale: Takeda suggests the guideline remove the reference to clinical as this may limit the types of studies that can be referenced as evidence of an instrument's psychometric properties. Psychometric measurement properties are often found in studies outside the clinical setting.	
Tanja Podkonjak, Takeda	15	Summary box	Takeda proposes that an additional bullet point be added to the summary box recommending the use of age-appropriate scales to highlight the importance of this factor.	It could be the same for other characteristics and it is impossible to provide an exhaustive list.
EFSPI	15	471-476	Interpretability of scales may vary in case categorical scales are transformed into continuous scales or vice versa. Whether a continuous or categorical scale is used to determine the endpoint of interest strongly depends on the underlying objective. One might be interested to explore the rate of patients who improved / maintained / worsened their symptoms compared to baseline (patient level objective), or the average change compared to baseline (within group perspective). It is difficult to argue which of the associated analysis complements the other as they address different questions. The risk of data dredging can be avoided by specifying key outcomes of interest in advance of the analysis which is typically done in the analysis plan. It is recommended to align key outcomes with the agencies in joint scientific advise meetings. The inflation of type I error rate might not be an exclusive problem here but also for subgroup analyses. **Current wording**: "While this expression of treatment effectiveness can enhance interpretability, this analysis on the categorical scale should complement the analysis on the continuous scale. In addition, to avoid the risk of data dredging and inflated type-1-error-rate, one measure of treatment effect should be pre-specified in the protocol and statistical analysis plan as a primary analysis (see the EUnetHTA 21 practical guideline "Applicability of evidence: practical guideline on multiplicity, subgroup, sensitivity and post-hoc analyses")." **Proposed wording**: "This expression of treatment effectiveness can enhance interpretability. Analysis on the categorical scale could complement the analysis on the continuous scale and vice versa. In addition, to avoid the risk of data dredging, one measure of treatment effect should be pre-specified in the protocol and statistical analysis plan as a primary analysis	This will be considered for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			(see the EUnetHTA 21 practical guideline "Applicability of evidence: practical guideline on multiplicity, subgroup, sensitivity and post-hoc analyses")."	
Norbert Gerbsch for IGES Institut GmbH and Health Econ AG	15	469-473	**Comment:** *"Here, relative treatment effectiveness can be expressed by a difference in the proportion of patients who have switched from categories (and/or by using an effect measure such as a risk ratio). While this expression of treatment effectiveness can enhance interpretability, this analysis on the categorical scale should complement the analysis on the continuous scale."* This suggests that a responder analysis is considered a complement to the analysis on the continuous scale. It should be clarified if this was the intended meaning, since a clinical relevance threshold is inherent to a responder analysis and therefore contains valuable additional information. If a continuous analysis is considered the primary analysis it should be clarified whether standardised mean differences such as Hedges' g will be required to assess clinical relevance. Otherwise, results on a continuous scale that are statistically significant might be based on differences, that do not involve clinical relevance.	It will be clarified for the next version of the draft.
Mihai Rotaru, EFPIA	15	450-451	Proposed change: *"Therefore, PROM translation must follow specific rules (**Wild et al, 2005), notably published guidelines (Epstein at al., 2015)** on transcultural adaptation ~~notably including~~ **and** a specific **cognitive debriefing** ~~validation phase~~ after translation**, or, ideally, the process recommended by the instrument developer**".* Rationale: Proposal to clarify that cognitive debriefing is typically required after translation of PROs rather than stating 'validation', hence our proposed change. Furthermore, there is a wide choice of translation and cross-cultural adaptation methodologies reported in the literature and different approaches generally achieve comparable results.[2] EFPIA recommends the statement be modified to reflect a broader set of methodologies. *Reference:*	These considerations will be clarified for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			1. Wild et al. (2005). Principles of Good Practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for translation and cultural adaptation. Value Health, 8(2): 94-104 *2. Cf. Epstein at al., A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. J Clin Epidemiol. 2015;68(4):435-41*	
Mihai Rotaru, EFPIA	15	473-474	Proposed change: "In addition, to avoid the risk of data dredging and inflated type-1-error-rate, one measure of treatment effect should be pre-specified in the protocol and statistical analysis plan as a primary analysis. ***However, endpoints that have not been pre-specified can also be accepted, if there are good reasons (e.g. endpoint methods estimation accepted by EMA, clinical plausibility, patient relevance, new composite endpoints).***"	This addition is not relevant to what is developed in the paragraph.
EFSPI	15	450-451	We propose to clarify that cognitive debriefing is typically required after translation of PROs rather than stating 'validation'. As well as this, an alternative reference is proposed: **Current wording**: "Therefore, PROM translation follows specific rules (transcultural adaptation [48]), notably including a specific validation phase after translation." **Proposed wording**: "Therefore, PROM translation follows specific rules (Wild et al 2005), notably including a specific cognitive debriefing phase after translation." Reference: Wild et al. (2005). Principles of Good Practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for translation and cultural adaptation. Value Health, 8(2): 94-104"	These considerations will be clarified for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Sebastian Werner vfa	15	473-474	*"In addition, to avoid the risk of data dredging and inflated type-1-error-rate, one measure of treatment effect should be pre-specified in the protocol and statistical analysis plan as a primary analysis."* The vfa recommends adding: *"if a valid or established MID exists for PRO, this should be used to avoid the risk of data dredging."*	We do not think this addition adds value to chat is developed in the paragraph.
Roche	15	450-451	Proposal to clarify that cognitive debriefing is typically required after translation of PROs rather than stating 'validation'. Additionally, an alternative reference is proposed: "Therefore, PROM translation follows specific rules **(Wild et al 2005)**, notably including a specific **cognitive debriefing** phase after translation." *Reference: Wild et al. (2005). Principles of Good Practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for translation and cultural adaptation. Value Health, 8(2): 94-104*	Duplicated comment.
Bayer	15	445-446	Comment: The recommendation "A measurement on a scale is valid and reliable only if it was computed using the measurement model as validated by the authors of the instrument." is considered too restrictive – particularly the part "validated by the authors of the instrument". It is common for researchers other than the original authors to assess structural validity of an existing PROM in new samples (e.g. a different patient population to the original development sample). This can lead to alternate measurement models (different factor structure or deletion of items), and therefore alternate scoring. Proposal: Recommend re-wording: *"A measurement on a scale is appropriate only if it was computed according to an evidence-based measurement model".*	This will be corrected for the next version of the draft.
Tanja Podkon	15	Summary box,	Current text: *"The two main properties of any measurement instruments are validity and*	We have not added responsiveness as it conceptually does not make much sense

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
jak, Takeda		bullet 1	*reliability."* Suggested wording: *"The ~~two~~ main properties of any measurement instruments are validity,* ***responsiveness*** *and reliability."* Rationale: Besides reliability and validity, responsiveness is also a key feature of measurement instrument, and we recommend it be added to the guideline in bullet 1, Summary Box on page 15.	to consider it a separate measurement property. Indeed, if an instrument is valid and is reliable, then change in scores, which is only the difference between two valid and reliable assessments, will be adequately measured. Therefore, this property is not distinguished in general metrology theory. Responsiveness is a concept that has mainly been developed in regard to the psychometric literature, but even within this field it has heavily been criticized. It has been retained in the Cosmin classification "for clarity", which means mainly because of usage. Last, studies that investigate "responsiveness" are usually non informative. Most of the time, they measure either effect sizes of change or if change can be considered statistically significant. But these empirical observations do not properly assess "responsiveness" as they do not confront the results against another criteria, which ideally would be the true change in scores (which is unknown in studies based on sample of patients).
Tanja Podkonjak, Takeda	15	Summary box, bullet 2	Current text: *"The validation of instruments is performed by specific clinical studies with appropriate design and statistical analyses."* Suggested wording: *"The validation of instruments ~~is~~* ***can be*** *performed by specific clinical or* ***observational*** *studies with appropriate design and statistical analyses."* Rationale: Validation of outcomes can also be done in observational studies in certain	These considerations will be clarifying for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			settings, and we recommend it be added to the guidance.	
Mihai Rotaru, EFPIA	15	Section 5.3 459-522	Some recommendations are in contrast to recent FDA guidance. Specifically, FDA is moving away from responder analysis, MID and MCID towards "meaningful within-patient change". FDA Guidance 4 discussion document states, "*In general, for COA-based endpoints FDA does not recommend a responder analysis endpoint or a percent change from baseline endpoint unless the targeted response is complete resolution of signs and symptoms. While percent change from baseline is popular in other contexts, the statistical measurement properties are poor*." The interpretability of scales section is generally focused on change in scores, and it is unclear what the recommendations are for PRO endpoints focused on maintenance and delays in deterioration. The discussion here may benefit from a statement indicating that a measure should demonstrate responsiveness to change in addition to evidence supporting reliability and validity. Several important references on interpretation are missing and the authors may want to consider them: • Cappelleri JC, Bushmakin AG. 2014. Interpretation of patient-reported outcomes. *Statistical Methods in Medical Research* 23:460-483. • Coon CD, Cappelleri JC. 2016. Interpreting change in scores on patient-reported outcome instruments. *Therapeutic Innovation & Regulatory Science* 50:22-29. • Coon CD, Cook KF. 2018. Moving from clinical significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. *Quality of Life Research* 27:33-40. • Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. 2007. Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal* 7:541-546. • King MT. 2011. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Review of Pharmacoeconomics and*	Considerations about responsiveness have already been addressed above. We agree with the remark on percent change from baseline, which is a distribution-based responder analysis, and we have pointed out issues about these types of RD within the guideline. We have not added recommendations about the interpretation of PROs by means of time to event analyses as we think there are conceptual issues to transform into "survival analyses" (which assumes the occurrence of an irreversible binary event in time) what is primarily a dynamic change in time on a continuous spectrum, and because these analyses of change in PRO scores are rarer. Provided references are already known by the authors and while they are relevant literature, we do not think citing all of them will add more value to the current guideline. "Meaningful within patient-change" is not conceptually very different from MID and MCID. This terminology is sometimes used because it was argued that the "D" in MID or MCID stands for between group differences rather that within group change. And also, because it was argued that finding "meaningful" change would be more valuable than finding "minimal" change (but this claim usually avoids the

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Outcomes Research 11:171-184. • McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. 2011. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Reviews of Pharmacoeconomics & Outcomes Research* 11:163–169. • Revicki D, Hays RD, Cella D, Sloan J. 2008. Recommended methods for determining responsiveness and minimally differences for patient-reported outcomes. *Journal of Clinical Epidemiology* 61:102-109.	fact that therefore there is a need for defining what is meaningful from a patient's perspective, which is a very polysemous term, while finding if the patient has at least felt the change is more straightforward). We have retained "MID" because this is the most prevalent term in the literature and because it would be unreasonable to expose all the debates about terminology in the guideline (minimal versus meaningful, difference versus change, within versus between groups…). But we will clarify that we use this term in relation to find a responder threshold that can be used to interpret within patient change over time. One the main recommendation of the FDA is to move away from using only one threshold. This in line with the guideline which develops either the possibility of deriving RD using multiple perspectives, triangulation for MIDs, or plotting results using cumulative distribution functions. Aside from minor differences in terminology, there are not major conceptual differences between what is proposed in this guideline and what is developed in FDA guidance.
Bayer	15	451	Proposal: Recommend adding a citation to useful guidance on transcultural adaptation: Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes	It will be considered if it is necessary for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			(PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005;8(2):94-104.	
Mihai Rotaru, EFPIA	15	Summary Box after Line 458	Proposed change: *"The ~~two~~* ***three*** *main properties of any measurement instruments are validity and reliability* ***and responsiveness to change over time****.* ***Responsiveness is also an important measurement property of COAs and refers to the ability of a COA to detect change over time in the construct to be measured[1]".*** Rationale: This section is missing the measurement property 'ability to detect change'. Hence our proposal to add this third key property generally considered in the PRO field and this is particularly critical for clinical trials. *Reference:* *1. Mokkink LB, Boers M, van der Vleuten CPM, et al. COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol 2020;20(1):293.*	Already addressed issue. See above.
Mihai Rotaru, EFPIA	15	Summary Box after Line 458 2nd bullet	*"The validation of instruments is performed by specific clinical studies with appropriate design and statistical analyses."* Comment: Validation of outcomes measures is not only done in clinical studies, but also done in observational studies. In addition, the instrument should be appropriate for the target population.	Clarification will be made for the next version of the draft.
Mihai Rotaru, EFPIA	15	Summary Box after Line 458 Last bullet	Proposed change: "*References, as provided by the HTD, allowing the access of the specific ~~(clinical)~~ studies assessing the measurement properties (and measurement model) of the instruments that are used."* Rationale: Remove "(clinical)" as this may limit the types of studies that can be referenced as	This will be corrected.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			evidence of an instrument's psychometric properties. Psychometric measurement properties are often found in studies outside the clinical setting. Furthermore, such evidence and studies may not have been published at time of EU HTA submission.	
Mihai Rotaru, EFPIA	15	Summary Box after Line 458 Additional bullet point	**Propose change 2:** EFPIA recommend the following bullet point be added: "*Where relevant, evidence of the transcultural adaptation methodology, as provided by the HTD*"	We will consider if this addition is necessary for the next version of the draft.
EFSPI	15	458	This section is missing the measurement property 'ability to detect change'. As this is a fundamental part of measure selection, we propose to add the following statement: "Responsiveness is also an important measurement property of COAs and refers to the ability of a COA to detect change over time in the construct to be measured (Mokkink et al. 2018, COSMIN)" Reference: Mokkink LB, Boers M, van der Vleuten CPM, et al. COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol 2020;20(1):293."	Already addressed issue. See above.
Roche	15	458	This section is missing the measurement property 'ability to detect change'. As this is a fundamental part of measure selection, we propose to add the following statement: "**Responsiveness is also an important measurement property of COAs and refers to the ability of a COA to detect change over time in the construct to be measured (Mokkink et al. 2018, COSMIN)"** *Reference: Mokkink LB, Boers M, van der Vleuten CPM, et al. COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol 2020;20(1):293.*	Already addressed issue. See above.
Prof.	15	From 459	Original wording:	As appraisal of the magnitude of treatment

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Matthias P. Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer nat, Laura Könenkamp, Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzens, M.Sc., Steven Krüger, M.Sc., SKC Beratungsgesellschaft mbH			"Interpretability can be defined as..." Comment: A clinically relevant threshold such as the minimal important difference (MID) needs to be specified by the individual MS, if considered relevant. The manufacturers need to know, which analyses need to be conducted in order to provide appropriate evidence. In general, since MIDs depend on the respective instrument, the validation, the characteristics of the patient population and the nature of clinical relevance, one specific MID should be valid across all member states.	effect if left at the discretion of member states, a unique consensual MID for all of them cannot be proposed.
Laurent Petit,	15	472	"*In addition, to avoid the risk of data dredging and inflated type-1-error-rate, one measure of treatment effect should be pre-specified in the protocol and*	We do not agree this sentence implies what is assumed in the comment. In only reports the risk of a factual methodological

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Leem			*statistical analysis plan as a primary analysis*" This sentence implies that any endpoints that would not be deemed statistically significant due to the statistical hierarchy would not be considered. - This method, although the most rigorous from a statistical point of view, sometimes fails to consider outcomes that cannot be reasonably covered within the hierarchical testing procedure. - Specifically**, HRQoL**, as PROs are generally not the primary endpoint, are considered as exploratory only. For example, France has a very low rate of consideration of HRQoL data, with HRQoL data being considered as exploratory only in about 80% of cases. - There is a need to reconcile PROs' importance and the concept of patient-centricity, and the consideration of HRQoL data when the latter do not satisfy the . **Suggestion : HRQoL data should be given consideration aside of the clinical efficacy data, even when outside of the statistical testing hierarchy.** The importance of assessing HRQoL data, especially in a patient-centric context, is crucial and the wording here should reflect the intention reflected in the rest of the guidance – with a focus on outcomes relevant to patients.	shortcoming and provides recommendation to avoid it in a specific situation (same outcome measured by more than one measure).
EFSPI	16	459-521	An overall a recommendation is missing about adequate determination of responders. Of note validated, established response thresholds are useful and valuable to enhance the transferability of risk/benefit assessments based on PRO measures to assess relative effectiveness for health technology assessment and to ensure consistent interpretation of PRO effects. A singular threshold of x%change of the continuous scale range for all instruments is incongruent with previously defined and scientifically established thresholds and is not well-suited for universal implementation. [Reference: Schlichting et al, Is IQWiG's 15% Threshold Universally Applicable in Assessing the Clinical Relevance of Patient-Reported	We did not really understand what the recommendation adds to the current content of the guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Outcomes Changes? An ISPOR Special Interest Group Report, Value in Health, Volume 25, Issue 9, 2022, Pages 1463-1468, ISSN 1098-3015, https://doi.org/10.1016/j.jval.2022.07.010.. Please add the following sentence: "In general, validated and established response thresholds should be considered when defining relevant individual response thresholds in terms of MID. Anchor based methods to determine MIDs utilizing patient-reported anchors are preferred. In absence of patient-reported anchors clinician reported anchors should be considered."	
Mihai Rotaru, EFPIA	16	477-520	Patient-reported outcome measures (PROMs) of an underlying continuous nature should be primarily analysed as continuous outcomes to detect treatment effect, and responder analyses should be reserved as secondary analyses for enhancing interpretability and for regulatory purposes of PROMs. Supportive References: 1. Collister D, Bangiwala S, Walsh M, Mian R, Lee SF, Furukawa TA, Guyatt G. Patient reported outcome measures in clinical trials should be initially analyzed as continuous outcomes for statistical significance and responder analyses should be reserved as secondary analyses. J Clin Epidemiol 2021;134:95-102. **2.** Cappelleri JC. Further reduction in statistical power for responder analysis of patient-reported outcomes with measurement error. Journal of Clinical Epidemiology. 2021; 140:200-201.	Clarifications will be made for the next version of the draft.
Mihai Rotaru, EFPIA	16	492-520	The authors may want to consider international guidance on this topic such as the recommendations included in the FDA Guidance PFDD Public Workshop Guidance 3 Discussion Document (fda.gov). FDA recommends the use of anchor-based methods supplemented with both empirical cumulative distribution function (eCDF) and probability density function (PDF). Please consider adding reference to the recent set of FDA guidance documents -- https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-	Use of CDF is encouraged within the guideline. We will consider if the addition of reference is necessary for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medica	
EFSPI	16	499-509	We acknowledge that distribution-based methods are informative to statistically characterize MIDs. However, for a relevant patient-centered outcome, responder definitions should be primarily consider anchor-based methods ideally utilizing patient-reported anchors. **Current wording**: "MIDs are also frequently estimated using distribution-based methods [51]. In contrast to anchor-based methods, only the overall variability in scores is used in distribution-based methods. Thus, they are criticized as they do not explicitly refer to the meaning of the change for patients (51). Two approaches are most common. The first is based on estimation of Cohen's d, which is computed by dividing the mean change in score by the standard deviation for the score at baseline. On the basis of results from experimental psychology, Cohen proposed a rule of thumb whereby d values of 0.2, 0.5 and 0.8 approximate effect sizes considered as small, moderate and large, respectively [54]. Although not initially developed for responder definitions, d values of 0.2 and 0.5 are still proposed as plausible MID values. A second approach relies on disentangling changes in score from measurement errors. For example, on the basis of empirical observations, 1 standard error of measurement has been suggested as a plausible MID [55]." **Proposed wording**: "MIDs are also frequently estimated using distribution-based methods [51]. In contrast to anchor-based methods, only the overall variability in scores is used in distribution-based methods. Thus, they are criticized as they do not explicitly refer to the meaning of the change for patients (51). Two approaches are most common. The first is based on estimation of Cohen's d. A second approach relies on disentangling changes in score from measurement errors [55]."	We agree. We will clarify this for the next version of the draft.
Sebastian Werner vfa	16	499-509	*"MIDs are also frequently estimated using distribution-based methods [51]. In contrast to anchor-based methods, only the overall variability in scores is used in distribution-based methods. Thus, they are criticized as they do not explicitly refer to the meaning of the change for patients (51). Two approaches are most common. The first is based on estimation of Cohen's d,*	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			which is computed by dividing the mean change in score by the standard deviation for the score at baseline. On the basis of results from experimental psychology, Cohen proposed a rule of thumb whereby d values of 0.2, 0.5 and 0.8 approximate effect sizes considered as small, moderate and large, respectively [54]. Although not initially developed for responder definitions, d values of 0.2 and 0.5 are still proposed as plausible MID values. A second approach relies on disentangling changes in score from measurement errors. For example, on the basis of empirical observations, 1 standard error of measurement has been suggested as a plausible MID [55].” Distribution-based methods are informative to statistically characterize MIDs. However, for a relevant patient-centered outcome, responder definitions should primarily consider anchor-based methods ideally utilizing patient-reported anchors. However, in the absence of patient-reported anchors, clinician-reported anchors should be acceptable. Responder definitions that are valid or established according to the generally accepted state of scientific knowledge and the international standards of evidence-based medicine should be accepted in the JCA report. These responder thresholds are useful to facilitate the alignment of regulatory and HTA assessments with consistent interpretation of PRO effects. A singular threshold of 15%-change of the continuous scale range for all instruments is incongruent with previously defined and scientifically established thresholds and is not well-suited for universal implementation [3] The vfa recommends adding as follows: *“Validated, and established response thresholds (incl. MID) should be considered in the requests of the Member states and accepted in the JCA. Anchor based methods to determine MIDs utilizing patient-reported anchors are preferred. In absence of patient-reported anchors clinician reported anchors should be considered.”* [3]: *Schlichting et al, Is IQWiG's 15% Threshold Universally Applicable in Assessing the Clinical Relevance of Patient-Reported Outcomes Changes? An ISPOR Special Interest Group Report, Value in Health, Volume 25, Issue 9, 2022, Pages 1463-1468,* *ISSN 1098-3015, https://doi.org/10.1016/j.jval.2022.07.010..*	

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Bayer	16	499-509	Proposal: To distinguish the suggestion that distribution-based methods can be used to estimate MIDs (or any threshold based on importance) from established methods to derive MIDs Comment: Distribution-based methods such as the SEM relate to the minimal detectable change: the smallest change that can be detected by the instrument beyond measurement error. This is a different concept to the MID and the two should not be conflated. See two papers by de Vet and colleagues for further discussion. Proposal: Recommend replacing the current paragraph with the following text: "If no anchor-based thresholds exist, distribution-based estimates can be used to support responder definitions. Distribution-based methods estimate the minimal detectable change: the smallest change that can be detected by the instrument beyond measurement error (de Vet & Terwee, 2010). This is a different concept to the MID and the two should not be conflated. Common distribution-based estimates are Cohen's d values of 0.2 or 0.5, or 1 standard error of measurement. A more direct estimate of the minimal detectable change is also available (de Vet et al., 2006)". (i) Henrica C.W. de Vet, Caroline B. Terwee, The minimal detectable change should not replace the minimal important difference, Journal of Clinical Epidemiology, 63(7), 804-805 (2010). de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes. 2006 Aug 22;4:54.	While we agree the two concepts can be conceived as different, and while we agree that anchor-based methods are the only ones that consider the patient's perspective, some authors have also argued that some distribution-based method could complement the use on anchor-based one to triangulate a plausible MID value. We will consider if more clarifications are needed for the next version of the draft.
Mihai Rotaru, EFPIA	16	485-491	Suggest adding the term "Meaningful Within-Patient Change (MWPC)" in the guideline. MWPC has been illustrated in the FDA Guidance PFDD Public Workshop Guidance 3 Discussion Document (fda.gov).	The authors are aware of the many, old, but still ongoing debates about the terminology within the literature about

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			The current statements do not distinguish the patient-level change (e.g., MWPC) and the group-level difference (e.g., mean difference between treatment groups). It is suggested to clarify the terms in patient-level or group-level when "MID" and "MCID" are used. The authors may want to consider that minimal important difference is a term that scientists have advised moving away from, replacing with clinically meaningful difference or clinically important difference. Some argue that MID and CMD are actually different values and CMD is the more relevant of the two.	interpretability (meaningful versus minimal, difference versus change, clinical or not clinical, between group versus within group…). Minimal Perceived Change has also been recently proposed, with the support of a conceptual model which for the first time defines this concept as a statistical parameter with a population-level definition. While all these debates are important, we have nevertheless retained the term MID as the guideline has a practical purpose first and MID is still the most prevalent term in the literature. We will nevertheless add a clarification that we use the term in the guideline in the context of deriving a RD threshold that helps to interpret within patient-change.
Sebastian Werner vfa	16	492-498	*"The methods that are usually considered the most appropriate for estimating MIDs are anchor-based methods, as they explicitly link a change in score to the patient's perception [51]. A change in score is linked to the response for a unique item: a patient global rating of change (PGRC). A PGRC is an overall assessment of a change compared to baseline performed by the patient. For instance, a PGRC can be phrased as follows: "Since the beginning of your treatment, overall, do you think your quality of life is now…". Proposed responses could be "a lot better", "a little better", "about the same", "a little worse" and "a lot worse"."* In the absence of patient-reported anchors, clinician-reported anchors should be used. The vfa recommends adding following sentence on line 498: *"When patient-reported anchors are not available clinician reported anchors are acceptable."*	We will consider this addition for the next version of the draft.
S.Walleser	16	485-491	Caution should be taken when using a MID or MCID in HTAs. Many of these are not properly validated and may be based on expert opinion. They should	We agree. This is why we ask for an access to the necessary bibliography that

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Autiero Medtronic			not be used to discount the importance of PROMS or patients' own experiences.	allow a proper assessment of the validity of proposed responder definitions.
Bayer	16	485-491	Statement: The text implies that MIDs can only be estimated through linking "to the subjective meaning of what is a relevant change according to patients." Comment: In practice, MIDs are often estimated using clinician-rated global impression scales, or performance status (Mouelhi et al., 2020). Observer-reported anchors are also used in paediatric populations (Silkey et al., 2022). (i) Mouelhi, Y., Jouve, E., Castelli, C. et al. How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. Health Qual Life Outcomes 18, 136 (2020). Silkey, M., Durán-Pacheco, G., Johnson, M. et al. The Autism Impact Measure (AIM): Meaningful Change Thresholds and Core Symptom Changes Over One Year from an Online Survey in the U.S. J Autism Dev Disord (2022).	We agree a responder definition can be derived using clinical anchors, which we will acknowledged within the guideline. But the very first definition of MCID in 1989 involves a subjective appraisal of the patient. Therefore, for clarity, we prefer to talk about clinical anchors as a way of deriving a "responder definition" in general (from the healthcare professional perspective), from "MID", a type of responder definition which focuses on the patient's perspective.
Bayer	16	485-491	Comment: The term MID is often used to refer to a threshold in terms of a between-group difference in scores (e.g. difference in mean change from baseline). Therefore, it is recommended to add a statement clarifying that a threshold for within-patient change is the intended interpretation and should be used for a responder definition. Proposal: Example statement: "Although the term MID has been used to describe a threshold to interpret between-group differences in scores (e.g. difference in mean change from baseline), the intention within this guideline is to refer to a threshold for within-patient change over time."	It will be added in the next version of the draft.
Bayer	16	492-498	Comment: In practice, multiple anchor-based estimates are commonly obtained and there is a need to perform 'triangulation'.	This is indeed regularly performed. We will consider if adding this precision is

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Proposal: Recommend adding this advice alongside a citation: "In practice, multiple anchor-based estimates are commonly obtained and there is a need to perform triangulation to derive a single threshold (or small range) [Trigg & Griffiths, 2021]" Trigg, A., Griffiths, P. Triangulation of multiple meaningful change thresholds for patient-reported outcome scores. Qual Life Res 30, 2755–2764 (2021).	necessary for the next version of the draft.
EFSPI	16	487-491	We suggest to clarify that the MID should be used to interpret important differences at the individual level - i.e. by analyzing the proportion of patients achieving the MID. This is in contrast to the common but problematic approach of comparing group-level mean differences to the MID - which does not capture the patient's perspective. **Current wording**: "Although this approach was initially developed for PROMs, it can be useful for other measurement instruments." **Proposed wording**: ""Although this approach was initially developed for PROMs, it can be useful for other measurement instruments. In general, MID should be used to interpret important differences at the individual level, i.e. by analyzing the proportion of patients achieving the MID."	We will consider clarifications for the next version of the draft.
Mihai Rotaru, EFPIA	16	486-488	Current wording*:* *"This approach is called the **minimal important difference (MID)** and can be defined as the minimal change in score perceived as an improvement or deterioration by the patient [50–52]."* Comment: In some diseases, especially rare disease the MID is not well established and/or no consensus exist. Patient reported outcome and/or patient perspective and clinical experts nominated to take part in the JCA can support any MID that shows an improvement. Therefore, EFPIA calls for inclusion of patient and clinical experts throughout the JCA production, including in review of the draft report to ensure meaningful interpretation of results where needed.	Patient's involvement during the JCA process is the concern of other guidelines.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			It is important to specify that MID is the change in score perceived as a meaningful or important improvement or deterioration by the patient.	
Mihai Rotaru, EFPIA	16	493-495	Current wording: *"A change in score is linked to the response for a unique item: a patient global rating of change (PGRC). A PGRC is an overall assessment of a change compared to baseline performed by the patient."* Rationale: This paragraph could be interpreted that PGRC (or patient's global impression of change (PGIC), as it is more commonly described) is the only appropriate anchor. Anchors can be clinical values or established clinical thresholds, or indeed, also outcomes reported by patients, but again, PGRC is not the only one. Further, if patient global impression-type of anchor is selected, the change in patient-global impression of severity (PGIS) is preferred because it avoids recall bias. It would be beneficial to provide an example of PGIS, if only one example is to be provided.	This sentence does not imply this is the only anchor possible, but this is the most consensual and used to derive a responder definition using the patient's own appraisal on change. We agree PGISs have sometimes been advocated for defining responder definition according to the patient's perspective, as proposed by the FDA as the end of the 2000s. However, we have not decided to describe this approach for multiple reasons. First, the uptake of this method is largely lower than the use of PGIC, and literature is scarcer about the validity and reliability of this method than the one based on PGIC. Second, while we do agree PGICs are prone to recall bias, there are conceptual issues with the use of PGISs that are frequently overlooked but are fundamental. First, they are a global measure of the targeted construct (e.g., HRQoL) at one time of measurement. In that regard, they can be considered as a redundant second assessment of the same outcome that the one that is measured by the instrument of interest but with less validity and reliability. It could therefore be argued that, while advocated as such, they are not anchor-based methods as they do not anchor the change in scores to another phenomenon. Rather,

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
				they repeat the measure of the same phenomenon. PGIC allows the measure of a different but related construct (i.e., the perceived change by the patient) to the construct of interest. Second, to define a RD using two PGISs, it is the researcher or healthcare professional who must define what is "change" (for example as a change in at least one unit between the two PGISs) before linking it to the change in scores of the instruments of interest. Therefore, the rule used for defining that there is change does not come from an assessment of the perception of the patients on that change. Thus, while they are advocated as such, RD computed by using two PGISs cannot be considered as MIDs according to the patient perspective.
Mihai Rotaru, EFPIA	16	510-512	Direct input from patients/qualitative work can also be used to establish the clinically meaningful change. The guidance would benefit from discussing the role of interviewing patients as a supplemental research tool to determine or confirm the meaningful change threshold should be included here.	Considerations that are exposed within the guideline are not meant to be exhaustive but are meant first for practical use to assessors. Therefore, we have focused content on the most prevalent methods. We will consider if expansion is wise for the next version of the draft.
Prof. Matthias P. Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer nat, Laura	16	477-479	Original wording: "In general, responder definition can be used to decide whether each patient has achieved a treatment benefit. This can be done either by assessing whether or not a patient reached a prespecified level of success, or by assessing whether the change in scores is as least equal to a pre-specified threshold" Comment: A responder definition should consider that the primary treatment goal in certain (mostly severe) indications is to maintain a former/current therapy response (e.g., cancer maintenance therapies) or to maintain a certain	Patient acceptable symptomatic state which is defined later in the guideline seems to be more appropriate in the context that is exposed in this comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Könenkamp, Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzens, M.Sc., Steven Krüger, M.Sc., SKC Beratungsgesellschaft mbH			disease/remission level or even just to slow down a negative trend of disease/health status (e.g., neurodegenerative diseases, late-stage/late-line cancer and other palliative settings). In conclusion, in certain indications, a missing or even slightly negative change in response scores might be the consequence of a positive therapeutic effect (compared to an otherwise strongly decreasing health status). A responder definition might therefore not necessarily include an “improvement”.	
Tanja Podkonjak, Takeda	16	494-496	Current text: *“A change in score is linked to the response for a unique item: a patient global rating of change (PGRC).”* Suggested wording: *“A change in score is linked to the response for a unique item: a patient global rating of change (PGRC). PGRC may also be referred to as Patient Global Impression of Change (PGIC).”* Rationale: As PGRC is commonly also referred to a PGIC in literature, Takeda recommends including both terms in the guideline.	It will be added in the next version of the draft.
Lieben	16	510 - 511	“MIDs are sometimes identified on the basis of expert opinion [51]. Such	It does not imply it has no value. It implies

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
hoff, BAH			MIDs are only a representation of what experts think about a change that patients consider significant." This sentence could give the impression that the opinion of experts has no value. MIDs or other response thresholds, based on experts' expertise should be appreciated. For that please change or better delete this sentence.	that this is not the same perspective than the patient's perspective. We will consider a deletion of the sentence.
Roche	16	493-494	Proposal to add in greater clarity regarding the anchor-based approach: "**A change in score on a COA is associated with an external measure known as an anchor. Commonly used anchors include global impression scales, usually consisting of a single item** (e.g., a patient global rating of change (**PGI-C**) **or severity (PGI-S)**. **These scales assess change or severity in a given construct (e.g., symptoms or quality of life) and the response categories (e.g., minimal improvement in the case of a PGI-C or a one or two category change in severity) can be used as an anchor to assess meaningful change on the outcome of interest.**	We will consider if such clarifications are needed.
EFSPI	16	477	**Current wording**: "In general, responder definition can be used to decide whether each patient has achieved a treatment benefit." Deciding whether a patient has achieved a treatment benefit requires knowledge of the counterfactual. We suggest to delete this sentence.	We will clarify content for the next version of the draft.
M. Ermisch – GKV-SV	16	485 ff	The information presented on MID lacks information that despite of the widespread use of MID to estimate treatment effects, the validity of MID is disputed. Systematic reviews of MIDs have shown that multiple different MIDs are available for individual scales, making the reliability of each of these MID questionable. Thus, responder analyses relying on these MIDs may not give valid results. As long as no scientific consensus on new MIDs can be established, the use of an alternative threshold for changes noticeable for patients should be used. We prefer the proposal IQWIG upholds in its general methods version 6.1, a threshold of 15 % (see: https://www.iqwig.de/methoden/general-methods_version-6-1.pdf, p. 173f). The guidance should be supplemented accordingly. Moreover, MIDs based on expert opinion alone (as outlines in line 510-11)	We agree there are still many methodological debates on how MIDs should be derived. But we do not think a systematic use for all MS of an arbitrary threshold can be considered consensual. Therefore, we prefer in this guideline the approach of presenting a more general account of the issue and we recommend HTDs to provide access to the bibliography allowing the assessment of the RDs that can be used in the submitted evidence.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			would not be acceptable at all.	
Natacha Bolanos, Lymphoma Coalition	16	485	The evaluation and use of patient reported outcome (PRO) measures requires detailed understanding of the meaning of the outcome of interest. Estimation of an MID and interpretation of clinical trial results that present patient important outcomes is as demanding as it is vital in informing the decision to recommend or not to recommend or approve a given intervention. Ideally we should get "gold standard" methodology of estimating the MID or achieving the meaningfulness of clinical trial results based on patient reported outcomes. The different approaches to estimating the MID will produce similar results. If they do not, this should be explicitly labelled. For instance, Fatigue is the most frequent symptom reported by patients with chronic illnesses including cancer. As a subjective experience, fatigue is commonly assessed with patient-reported outcome measures (PROMs). Currently, there are more than 40 generic and disease-specific PROMs for assessing fatigue in use today. The interpretation of changes in PROM scores may be enhanced by estimates of the so-called minimal important difference (MID), but Magnitudes of published MIDs for fatigue PROMs vary considerably. Consistent information about the derivation of fatigue MIDs is needed to evaluate their applicability and suitability for use in clinical practice and research.	We agree. But we are not responsible for the fact that the issue of interpretability of outcome measurement instruments is still an area where methodological developments are needed.
Laurent Petit, Leem	16	486	*MIDS :* As though it is relevant and important to introduce the concept of Minimal importance difference to ensure a patient-centred evaluation, there is a need to move a way from a hard scale and table that might set the MID. **Suggestion :** **Elements should be considered in the setting of MIDs depending on each clinical trial and appreciative of several key concepts :** - **Disease stage**	We will consider if clarifications on factors that can lead to variation in MID values will be needed for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			- **Duration of clinical trial (as MID must be set in consideration with time necessary to reach it)** *"The patient's perspective is frequently used by linking a change in score to the subjective meaning of what is a relevant change according to patients. This approach is called the minimal important difference (MID) and can be defined as the minimal change in score perceived as an improvement or deterioration by the patient"*	
Roche	16	498	Proposal to clarify at the end of this paragraph that meaningful change can be assessed between groups and at the individual patient level and that these estimates are not interchangeable: **"Commonly, trial endpoints assess treatment efficacy via the difference in change score between treatment arms (e.g., LS Mean difference via ANCOVA or MMRM). Estimates of meaningful individual change (e.g., responder/progressor definitions) are inappropriate for direct use to interpret the magnitude of the score change difference (i.e., a meaningful change score should not be used to interpret the difference between two change scores). Instead, the proportion of patients meeting the change score threshold or greater can be compared across treatment arms to provide context on the meaningfulness of treatment with the investigational drug."**	Already addressed issue.
HTAi PCIG	16	Line 485 and Line 510	It is laudable that the authors identify that a difference is only meaningful if it is meaningful to patients. However, further below (line 510), the document describes *"MIDs are sometimes identified on the basis of expert opinion [51]. Such MIDs are only a representation of what experts think about a change that patients consider significant."* Here, it should be clearly stated what is meant with 'experts' and that the value of a clinician- or KOL-determined MID is less than a Patient determined MID.	This sentence will be deleted as it does not hold much value for the purpose of this document.
AIM – International	17		Some parts of the sentences of the bullet points are missing ?	Thank you we will correct the sentence

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Association of Mutual Benefit Societies				
Mihai Rotaru, EFPIA	17	Summary Box after Line 521	Requirement for JCA reporting – 3rd bullet Proposed change: "*References, as provided by the HTD, to allow full access to the* ***bibliography*** ~~*literature*~~ *justifying the responder definitions used. "* Rationale: We would like to use the word bibliography, to avoid any confusion. Indeed, physical or electronic copies of the papers often cannot be provided due to copyright issues on sharing papers	It will be modified.
Tanja Podkonjak, Takeda	17	Summary box	Takeda suggests the following additional bullet be added: *'A classification of response based on the degree of response may be useful in some circumstances instead of a dichotomization of the outcome."*	This approach is described at the beginning of the section.
Mihai Rotaru, EFPIA	17	Summary Box after Line 521	Suggest adding in the following Summary bullet point: *"In some cases, a classification of response by strength/degree of response might be useful instead of a dichotomization of the outcome."*	This approach is described at the beginning of the section.
Ioanna Psalti EUEYE	**6 & 12**	**160-161; & 367**	**Disability/impairment distinction** The document uses the terms impairment, disability and incapacity and it is not clear whether these terms are interchangeable given the social and medical dimensions involved in the terms. The social model of disability distinguishes between impairment and disability, identifying the latter as a disadvantage that stems from a lack of fit between a body and its social environment. This distinction has a	These terms are only used in the guideline as possible examples of outcomes that can be assessed. Their use does not imply they are used interchangeably neither they endorse which should be best assessed from an HTA perspective.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			profound impact in chronic disease with increasing impairment as their outcome. Measures used in HTA should focus on impairment/disability/incapacity that cannot be remedied by social change and they should be able to differentiate between impairment/disability/incapacity caused by disease progression and that caused by treatment e.g. radiation in cancer. It may be important that this is highlighted in the document particularly when assessing a health technology on the basis of outcomes such as anxiety and depression which may be linked to social factors rather than impact of a therapy.	
Mihai Rotaru, EFPIA	21	Appendix A	Comment: In oncology, other surrogates have been used in recent clinical trials, for example minimal residual disease (MRD) is becoming important and common for instance for multiple myeloma and being investigated in other tumours. As an example, multiple myeloma is a haematologic malignancy characterised by the accumulation of malignant plasma cells, usually within the bone marrow. Long considered incurable, significant advances in the treatment of MM through the introduction of new therapies have contributed to an increase in patient survival rates from three years to more than five years. However, as a result, demonstrating OS benefit became a challenging objective for clinical trials, and most investigators, as well as health agencies, now focus on other endpoints, such as PFS or CR rates. Even these measures still present challenges in terms of the time required to collect and assess the data, and health agencies are currently examining MRD as an earlier endpoint in clinical trials. Recent publications have supported the position that achieving MRD negativity after treatment is a strong predictor for both longer PFS and OS. (1,2,3). The MRD in multiple myeloma is an example of the value of novel endpoint to measure efficacy in a context where OS data is not available. Similar novel endpoints are being developed for other diseases in oncology and ATMPs, areas where data on final endpoints such as OS are not feasible at time of assessment. These novel endpoints are important to keep pace with the development of science. EFPIA would	Thank you for the information

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			like to note that endpoints used in a clinical trial, particularly novel endpoints, are not decided in isolation by the HTD, they are discussed with clinical experts and agreed with the Regulator to ensure they are robust and clinically valid at the time of study design. Note: Minimal residual disease is the name given to small numbers of cells that remain in the person during treatment, or after treatment when the patient is in remission. An MRD negative result means that no disease was detected after treatment (4) Reference: 1. Perrot A, Lauwers-Cances V, Corre J, et al. Minimal residual disease negativity using deep sequencing is a major prognostic factor in multiple myeloma. Blood. 2018;132(23):2456-2464. doi:10.1182/blood2018-06-858613 2. Nikhil C. Munshi, Herve Avet-Loiseau, Kenneth C. Anderson, Paola Neri, Bruno Paiva, Mehmet Samur, Meletios Dimopoulos, Margarita Kulakova, Annette Lam, Mahmoud Hashim, Jianming He, Bart Heeg, Jon Ukropec, Jessica Vermeulen, Sarah Cote, Nizar Bahlis; A large meta-analysis establishes the role of MRD negativity in long-term survival outcomes in patients with multiple myeloma. Blood Adv 2020; 4 (23): 5988–5999. 3. Avet-Loiseau H, Ludwig H, Landgren O, et al. Minimal Residual Disease Status as a Surrogate Endpoint for Progression-free Survival in Newly Diagnosed Multiple Myeloma Studies: A Meta-analysis. Clin Lymphoma Myeloma Leuk. 2020;20(1):e30-e37. doi:10.1016/j.clml.2019.09.622 4. Minimum Residual Disease Fact Sheet, Leukemia & Lymphoma Society, 2019	
Mihai Rotaru, EFPIA	21	Appendix A Line 688-691	Proposed change: *"PFS is measured by censoring patients who are still alive* **and progression-free** *at the time of evaluation or those who were lost to follow up and thus the data are available earlier, within the timeframe of the trial."* Rationale: PFS is often an objective measure and in many studies validated by a blinded independent clinical review board, in addition to the clinical investigator or treating	Thank you for your comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			clinicians. PFS should, as with other clinical endpoints, treated as a clinically important endpoint in its own right, rather than only as surrogate for OS (esp. in earlier line therapies). In addition to informing prognosis for a patient depending on the speed of relapse or progression (i.e. refractory patients), a progression event is also used to stop existing treatment, change courses of treatment, inform the future treatment strategy all which have an impact on the patient and the health care system – these are therefore important outcomes which should be considered in the totality of evidence for a JCA as they are relevant for decision marking. Furthermore, there is a psychological impact of progression and relapse on patient which is also an important consideration when evaluating the overall effectiveness of a new technology compared to the existing standard of care. EFPIA requests the guideline be updated to include PFS as a stand-alone clinically relevant endpoint for oncology. In some circumstances, PFS may also be a relevant measure for some adjuvant settings, where there is remaining disease or partial response to adjuvant therapies, and we would recommend the guidance acknowledges this.	
Mihai Rotaru, EFPIA	21	Appendix A Line 703-704	*"The definition of "disease-free interval" is not always clear and the validity of an incidental finding of cancer regardless of symptoms has been questioned (57)"* We consider this statement to be misleading. The paper referenced (Clinical endpoints in oncology – a primer) acknowledges the above point that the validity of an incidental finding of cancer regardless of symptoms has been questioned, but also states that "DFS's use as a primary clinical endpoint in the setting of adjuvant therapies is supported by the fact that a large portion of patients will have cancer symptoms at the onset of disease recurrence" [57] which, for balance, we feel should also be included here.	We will consider if clarifications are needed.
Mihai Rotaru, EFPIA	21	Appendix A Line 711	Proposed change: *"For each unique tumour type, treatment **class**, and stage of disease".* Rationale: Avoiding the implication that surrogacy validation should be performed for every individual therapy.	We will consider if this precision is necessary.
Ioanna	**21**		The outcomes on Appendix A, page 21 are largely binary, ie	The appendix is meant to propose

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Psalti EUEYE			time to death, event free survival, etc and do not easily translate into ophthalmology where much of the ongoing treatment is to slow a disease without such categorical endpoints. Moreover interventions in Ophthalmology are also much less invasive than those in cancer with numerous rounds of chemo; time to next intervention of one month for the next injection for an eye condition does not equilibrate with one month in oncology for next round of chemo. Therefore these outcomes may be useful in guiding treatments in intraocular tumours but they are not applicable to the ophthalmology. Consider the proposal further below.	definitions of usual outcomes in oncology only as oncology medicines will be the first class of treatment assessed at the beginning of JCAs in 2025. It is not meant to be a list of outcomes for all medical areas. The title of the appendix specifies it.
Mihai Rotaru, EFPIA	21	Appendix A Lines 677-682	Proposed change: *"As in other treatment areas the OS has been regarded as the final patient-centred outcome in oncology (57). Improvement in OS clearly demonstrate clinical benefit which is meaningful to the patients. However, measuring OS often requires a large number of patients and long follow-ups. Long-term survival OS-data for the technology under assessment may be influenced by treatment given in further steps, sequential use of other agents, or even cross-over treatments, making it difficult to attribute the OS result to a specific medical intervention* ***(see for example, Lux et al. 2021). Additionally, it is becoming evident that other endpoints add valuable information about quality of life and treatment failure as their use is becoming increasingly prevalent in oncology clinical trials (57)"*** Comment: This introduction outlines how OS has been the gold-standard and goes on to describe the current practical challenges with demonstrating OS. However, what is missing from this introduction is the second challenge with reliance on OS, which is clearly stated in a paper referenced throughout this appendix ("Clinical endpoints in oncology – a primer"), that: "Additionally, it is becoming evident that other endpoints add valuable information about quality of life and treatment failure as their use is becoming increasingly prevalent in oncology clinical trials" [57] – EFPIA requests this text be added to the statement above.	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			As with our General comment (nos 4), the discussion of each clinical endpoint calls out surrogacy to OS, but does not refer to the ability of them to provide additional insights into a wider definition of patient-centred endpoints, including HRQoL etc… and informed clinician-patient decision-making. This is also discussed in "Clinical endpoints in oncology – a primer" (https://pubmed.ncbi.nlm.nih.gov/33948349). Reference: Lux MP, Ciani O, Dunlop WCN, Ferris A, Friedlander M. The Impasse on Overall Survival in Oncology Reimbursement Decision-Making: How Can We Resolve This? Cancer Manag Res. 2021 Nov 10;13:8457-8471. doi: 10.2147/CMAR.S328058.	
EFSPI	21	706-711	Definition of EFS is potentially misleading. Different trials use different definitions of EFS, most notably to sometimes account for the fact that some patients will be recorded as having an EFS event at Day 1. We would suggest amending this definition to account for the fact that EFS may sometimes adopt a disease-specific definition (e.g., failure to achieve remission by time x counts as an EFS event at Day 1), which can make this outcome very different to PFS or DFS.	We will consider it for the next version of the draft.
GSK	21	677-682	*"As in other treatment areas the OS has been regarded as the final patient-centred outcome in oncology (57). Improvement in OS clearly demonstrate clinical benefit which is meaningful to the patients. However, measuring OS often requires a large number of patients and long follow-ups. Long-term survival OS-data for the technology under assessment may be influenced by treatment given in further steps, sequential use of other agents, or even cross-over treatments, making it difficult to attribute the OS result to a specific medical intervention"* This introduction outlines how OS has been the gold-standard and goes on to describe the current practical challenges with demonstrating OS. However, what is missing from this introduction is the second challenge with reliance on OS, which is clearly stated in a paper referenced throughout this appendix ("Clinical endpoints in oncology – a primer"), that: "Additionally, it is becoming evident that other endpoints add valuable information about quality of life and treatment failure as their use is becoming	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			increasingly prevalent in oncology clinical trials" [57] We feel this is a key point that should be included in the appendix. The discussion of each endpoint calls out surrogacy to OS, but does not refer to the ability of TTP, DCR and CBR to provide additional insights into the quality of a patient's life (again, something highlighted in "Clinical endpoints in oncology – a primer").	
GSK	21	691-694	*"PFS seems to be frequently used surrogate endpoint in oncology since it can be reported within a shorter time of follow-up and the results may be obtained with a lower number of patients. However, the correlation between PFS and OS seems to differ across cancer types and therapy lines"* There is no reference in this document of the acceptability of PFS in the adjuvant setting as a measure in its own right. While the warning against the use of PFS as a surrogate for OS where correlation is not confirmed is important, we feel the acceptability of PFS in the adjuvant setting should be explicitly acknowledged, as in other EUnetHTA guidance. For example, both "Endpoints used in Relative Effectiveness Assessment: Surrogate Endpoints" and "Endpoints used in Relative Effectiveness Assessment: Clinical Endpoints" state the following about PFS: "The acceptability of progression free survival has not the same impact in adjuvant or metastatic disease. In the adjuvant setting, PFS appears acceptable; in the metastatic setting, PFS alone is insufficient; it might be considered if coupled with quality of life assessment and survival data, the maturity of which will be considered on the case by case basis." [14]	We will consider if this precision is necessary for the next version of the draft.
Mihai Rotaru, EFPIA	21	693-695	**Appendix – Progression-free survival** Proposed change: *"The correlation between PFS and OS is not always is confirmed by the final results~~, especially in studies of targeted therapy or immunologic agents (59)~~.* ***This may be due to the unique mechanism of action of targeted and immunologic agents (e.g., relatively low PFS improvements, but dramatic improvements in OS due to long periods of post-progression survival) (59, 60, 61).***"[1,2]	We do not think this suggestion is useful.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			Rationale: EFPIA is concerned that the current statement may potentially dismiss the value of PFS as an outcome since it is an important endpoint for both clinical research (e.g., commonly used as a primary endpoint), patient-centred outcome in oncology, and clinician-patient decision-making. Furthermore, particularly in this context for immunologic medicines, a lack of correlation should not be considered problematic because the immunological reaction is triggered by the treatment, the first progression can occur early, which ultimately leads to better OS in the long-term. As such, we have proposed this additionally wording given this therapy specific nuance that may be incorrectly interpretated as being problematic for the purposes of PFS.[1,2] References: 1. Tap WD, Jones RL, Van Tine BA,. et al. Olaratumab and doxorubicin versus doxorubicin alone for treatment of soft-tissue sarcoma: an open-label phase 1b and randomised phase 2 trial. Lancet. 2016;388:488–97. 2. Zer A, Prince RM, Amir E,. Et al. Evolution of Randomized Trials in Advanced/Metastatic Soft Tissue Sarcoma: End Point Selection, Surrogacy, and Quality of Reporting. J Clin Oncol. 2016;34:1469–75.	
GSK	21	701-703	*"DFS has been used as a surrogate outcome for OS in clinical trials for stage III colon cancer, in an adjuvant setting in lung cancer, and in breast cancer"* This has been taken almost word for word from "Clinical trials in oncology – a primer". Given that guidance D4.4 frequently refers to only using surrogates where there is strong evidence of correlation, we suggest rewording this statement to reflect the wording of the paper referenced – see wording **in bold**: "DFS has been used as a **strong** surrogate endpoint for OS in clinical trials for stage III colon cancer, in an adjuvant setting in lung cancer, and in breast cancer" Otherwise, the reader can misinterpret the statement to mean that DFS has been used as a surrogate inappropriately.	We do not think this ,addition is useful.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Prof. Matthias P. Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer nat, Laura Könenkamp, Dr. rer. nat., Dominik Müller, Dr. rer. nat., Sebastian Vinzens, M.Sc., Steven Krüger, M.Sc., SKC Beratungsgesellschaft mbH	21	693-695	Original wording: “However, the correlation between PFS and OS seems to differ across cancer types and therapy lines (58). The correlation between PFS and OS not always is confirmed by the final results, especially in studies of targeted therapy or immunologic agents (59).” Comment: As stated, the correlation between PFS and OS seems to differ across cancer types and therapy lines. Therefore, based on the specific indication and/or therapy line, a correlation between PFS and OS might be given and robust. In conclusion, a correlation of PFS and OS and the use of PFS as an early measurable surrogate for OS should not be excluded generally. Furthermore, PFS has high value as an efficacy outcome measure independent from a potential (indication- / therapy-line specific) correlation with OS. In contrast to OS, PFS is not affected by (potentially several) post-progression therapies outside of the clinical setting and enables to draw direct conclusions on immediate efficacy of the tested therapeutics. Not only the total length of survival, but also the duration of progression-free survival is of high patient relevance, due to the direct impact of tumor progression on symptoms, psychology, HrQOL and the postponing of following (mostly burdensome) chemotherapies or other cancer therapies, that are given in response to tumor progression. Further, in some cancer therapies, such as maintenance therapies, it is a defined primary treatment goal to prevent (new) tumor progression after previous response to an earlier therapy. Progression-free survival adequately captures such highly important aspects of cancer therapy and should be considered patient relevant.	Already addressed issue.
Mihai	21	683-684	Proposed change:	Already addressed issue.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Rotaru, EFPIA			*"In oncology most often reported disease related outcomes are* **objective response rate (ORR)**, *progression free survival (PFS)* ~~*as surrogate for OS*~~*, event free survival (EFS), or disease-free survival (DFS)."* Rational: It is unclear why PFS has been labelled at a surrogate for OS, whereas this distinction has not been made for EFS and DFS. EFPIA believe PFS, along with EFS and DFS, is an important clinical and patient-relevant outcome, irrespective of it being a potential surrogate for OS. We are aware of several recent studies that report patients have a strong preference for longer periods of time when the disease is not progressing, even when OS may not be extended[1,2]. In addition, we believe objective response rate (ORR) should be added to the list of outcomes. ORR is commonly used outcome in oncology trials and is used in clinical practice to help inform a patient's long-term care and expected outcomes[3]. References: 1. MacEwan J, Doctor J, Mulligan K et al. The value of progression-free survival in metastatic breast cancer: results from a survey of patients and providers. MDM Policy and Practice 2019; 1-14. 2. Mertz S, Benjamin C, Girvalaki C et al. Progression-free survival and quality of life in metastatic breast cancer: the patient perspective. Breast 2022; 65: 84-90. 3. Goring S, Varol N, Waser N et al. Correlations between objective response rate and survival-based endpoints in first-line advanced non-small cell lung cancer: a systematic review and meta-analysis. Lung Cancer 2022; 170: 122-132.	
GSK	21	703-704	*"The definition of "disease-free interval" is not always clear and the validity of an incidental finding of cancer regardless of symptoms has been questioned (57)"* We respectfully consider this statement to be misleading. The paper referenced (Clinical endpoints in oncology – a primer) acknowledges the above point that the validity of an incidental finding of cancer regardless of symptoms has been questioned, but also states that "DFS's use as a primary clinical endpoint in the setting of adjuvant therapies is supported by the fact that a large portion of patients will have cancer symptoms at the onset of	Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
			disease recurrence" [57] which, for balance, we feel should also be included here.	
Tanja Podkonjak, Takeda	21	691-692	Current text: *"PFS seems to be frequently used surrogate endpoint in oncology since it can be reported within a shorter time of follow-up and the results may be obtained with a lower number of patients."* Proposed revision: ~~*"PFS seems to be frequently used surrogate endpoint in oncology since it can be reported within a shorter time of follow-up and the results may be obtained with a lower number of patients."*~~ Rationale: This statement is misleading, speculative and does not provide any supporting evidence. The choice of endpoints in a clinical trial, especially primary endpoints (often PFS in oncology), is based on discussions and agreement with regulators, clinical experts and patient experts with clinical development teams and is highly complex. D4.4 Endpoints is a critically important guideline which will impact the future EU Joint Clinical Assessments (JCA) and potentially availability of innovation to patients in Europe, the recommendations provided should be based on evidence. The current statement is unfounded without any references cites to support this claim and we request it be removed from the guidance.	The sentence does not imply these are the only factors that influence the choice of PFS.
M. Ermisch – GKV-SV	21	688 ff	The presentation for PFS as a putative surrogate for OS is far too positive. The number of diseases and disease states for which its surrogacy has been verified is very limited. Thus, a line should be added stating: "Thus, the surrogacy of PFS is disputed and needs to be justified by HTD: Generally, its acceptance in national decision making is limited."	These discussions are already addressed in the section about surrogate outcomes.
M. Ermisch – GKV-SV	21	696 ff	While it is true that TTP requires clear definition, this is no speciality but applicable for any other endpoint, too. However, the surrogacy of TTP is also disputed, if given at all. This should be mentioned accordingly. "TTP is usually not a patient relevant outcome; thus, its acceptance in national decision making is very limited."	We will consider if this addition is necessary for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
M. Ermisch – GKV-SV	21	700 ff.	DFS: The sentence „*DFS has been used as a surrogate outcome for OS in clinical trials for stage III colon cancer, in an adjuvant setting in 702 lung cancer, and in breast cancer*.“ should be deleted. It is neither a complete list nor is it clear to what extent the uncertainties mentioned below, such as a clear definition, were adequate in these examples	The sentence does not imply its aim is to provide a definitive list but just to outline some frequent usage.
GSK	21	711	Suggested rewording in bold to avoid the implication surrogacy validation should be performed for every individual therapy: “for each unique tumour type, treatment **class**, and stage of disease”.	Duplicated comment.
M. Ermisch – GKV-SV	21	712 ff	Objective response rate may be a suitable marker for anti-tumour activity, but is not a patient-relevant outcome per se. It may be a marker for morbidity, provided that response comes with a relief in symptoms. RECIST provides information on tumour response by imaging alone and can, thus, not provide a patient relevant outcome. Proposal for an addition: “ORR usually is not an acceptable outcome for HTA; RECIST criteria provide objective information on tumour response but lack a proof for patient relevancy. Thus, acceptance in national decision making is limited.”	Thank you. We will consider your suggestion next draft
Prof. Matthias P. Schönermark, M.D., Ph.D., Ingo Hantke, Dr. rer nat, Laura Könenkamp, Dr. rer. nat., Dominik Müller,	21	712 and following	Original wording: “Objective response rate (ORR) is a measure of antitumor activity and defines a proportion of patients that respond either partially or fully to the therapy according to a predefined set of response criteria….” Comment (also see comment for line 477 – 479, regarding response criteria and comment for line 693 – 695, regarding PFS): ORR has high value as an efficacy outcome measure and should be considered patient relevant. Tumor response is of high patient relevance, due to the direct impact of tumor progression on symptoms, psychology, HrQoL and the decision (postponing or initiation) of following (mostly burdensome) chemotherapies or other cancer therapies, that are given in response to tumor progression. The tumor response adequately captures these highly important aspects of cancer therapy. The monitoring of success of tumor-deceleration per RECIST-criteria is highly relevant since it allows to adapt the therapy regime before the condition gets too burdensome or the tumor progress gets too worse to be able to react effectively. Therefore, in therapeutic reality it is not	Thank you for your comments. ORR is usually not considered as patient relevant outcome in HTA although it is useful for decision making on further treatment and thus relevant for the patient. See comment above.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	HOG answer
Dr. rer. nat., Sebastian Vinzens, M.Sc., Steven Krüger, M.Sc., SKC Beratungsgesellschaft mbH			applicable just to wait until a symptom strongly deteriorates or a new one occurs. For this reason, not only endpoints that are immediately describing symptoms should be considered patient relevant. Also, in contrast to OS, and equivalent to PFS, ORR is not affected by (potentially several) post-progression therapies outside of the clinical setting and enables to draw direct conclusions on immediate efficacy of the tested therapeutics.	
Dr. Martin Danner BAG SELBSTHILFE Germany	26	280	After COS should be added: “The lack of patent-reported outcomes always has to discussed with the involved patient experts to see if the use of COS is accepted” Comment: The draft notices the missing of patient reported outcomes in many COS so this has to be enlightened obligatory in the joint assessment / the joint consultation.	Thank you for your input

EUnetHTA 21 Public Consultation Comments and Responses
Of D4.4 - Endpoints

Comments received from stakeholders outside of EU/EEA countries

General comment from the Hands of Groups: We thank all the stakeholders outside of EU/EEA countries for their insightful comments, particularly the ISPOR. However, this practical guideline on methodological aspects was the one for which we have received the greatest number of comments from stakeholders within EU/EEA. As EUnetHTA 21 guidelines are a production that is a contract between EU Commission and EUnetHTA, the HOG did have to comply with a precise timeframe for guideline development. Therefore, we have prioritized answers to stakeholders related to EU/EEA countries. All comments below have been acknowledged though, and we will better consider it for subsequent steps of the process of developing JCAs framework under the HTAR.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
ISPOR – The Professional Society for Health Economics and Outcomes Research	General		It is suggested to look beyond the references for standardization purposes. The definitions could be a bit more clear by using some more up to date documents, and citation of regulatory documents or papers from regulatory bodies (e.g., EMA reflection papers, FDA guidance) would support harmonization across the sector. Further to this point, and to the goal of harmonization and standardization, it would also be helpful if definitions for key terms, such as those around COAs, came from recently published documents or drew upon commonly used/widely cited documents from professional societies. One very relevant document is the recent FDA draft guidance on Clinical Outcome Assessments: https://www.fda.gov/about-fda/clinical-outcome-assessment-coa-frequently-asked-questions		Thank you for the information
ISPOR – The Professional Society for Health Economics	General		The guideline in its current version summarizes general concepts but still leaves large room for interpretation and specific application of concepts between the member states (MS): Based on scientific rationale there should be more harmonization between MS regarding the following		No, this is not required by HTAR

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
and Outcomes Research			aspects: - Clinical relevance - Acceptance of surrogate endpoints - Acceptance of Responder definitions		
ISPOR – The Professional Society for Health Economics and Outcomes Research	General		Much of the guidance is stated in fairly general terms and refers to good clinical practice or good statistical practice, which is per se meaningful as some decisions have to be made on a case to-case basis. However, a key concept of the entire JCA is also to have sufficient certainty about the acceptance of the approaches used by the HTD and on the implementation of good scientific practice in the specific situation of interest. Therefore, the key parameters for implementing outcomes, surrogate validation, validity, reliability and responder definitions should be determined in advance by a close exchange between assessors and HTD – on the basis of state-of-the-art scientific methods. The outcome of this exchange should be binding in the sense that the agreed methods are accepted in the final assessment		See previous point
ISPOR – The Professional Society for Health Economics and Outcomes Research	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.		As already stated, this guideline is not intended to never be updated
ISPOR – The Professional Society for Health Economics and Outcomes	General		It would be helpful to have a process map clarifying roles of EMA, EUnetHTA, HTA, HTD, and MS in defining, requesting, submitting, and assessing COA in the JCA at the beginning.		Indeed, it would be helpful. This is outside the scope for this guideline, but maybe could be a project for ISPOR?

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Research					
ISPOR – The Professional Society for Health Economics and Outcomes Research	General		Since there seems to be some variation in volume among the items, we suggest that the existing guidelines be used as a reference and that a little more balance be achieved.		Thank you for your comment
ISPOR – The Professional Society for Health Economics and Outcomes Research	General		The document could be more clearly divided into recommendations for Member States and recommendations for how to write the JCA report. We would also welcome an overall aims description for assessing the outcomes and impact of pharmaceutical products. Pharmaceutical products are not only with the aim to prolong life, but also for patients to be healthy and productive, to enable work participation, to enable living in own home etc. Especially given the demographic challenges we are facing in most countries. To keep chasing mortality as main outcome seems outdated to me. I do not think this guidance is reflecting the needs in the health care sector and in societies today. In addition, the use of proxies/surrogate endpoints should be encouraged where these can be used to (adequately) predict other desired outcomes. This should be so for ethical and economic reasons. I think the document should consider the totality of the health care environment in the future.		The outcome chosen at EU level through the consolidated PICO should reflect MS needs. As stated in the D4.2 Scoping Process, outcomes of interest are to be decided at MS level only. There is not requirement in the HTAR to endorse or encourage some outcome in advance.
Pooyeh Graili Quality HTA	General		The content is helpful for the user if the outline and the content be organized, to clarify the text by using simple, understandable language to be	x	Out of scope (see D4.6 Validity of clinical study)

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			easily understandable for users. The content needs some modifications to be aligned with definitions and requirements in practice. A piece of information missing in this document is RWE and the outcomes we can achieve with RWD, although the outcomes gained by digital health were pointed out. There are different sources of RWD which result in a range of RWE and is important to be considered. Acronyms make sense only when an expression is used more than once in the text. HaDEA needs to be corrected as the acronym of European Health and Digital Executive Agency I would suggest using 'r' instead of 'R' as the acronym of HTA Regulation Couldn't 'Health technology developer' have no acronym and instead, the word 'developer' be used to refer to Health Technology Developer? Why do we need an acronym for 'Medical Dictionary for Regulatory Activities', 'Minimal clinically important difference', 'Medical Outcome Study', 'Patient-acceptable symptomatic state', 'SUSAR', and 'WHO-ART'? These have not been repeated more than once or twice in the text.	X X X x	
ISPOR – The	17-21	6 References (Lines	The link to the EUnetHTA endpoints guideline (ref no. 14)		Thank you. We will make necessary

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Professional Society for Health Economics and Outcomes Research		522-674)	is not valid. It should be https://www.eunethta.eu/wp-content/uploads/2018/01/Endpoints-used-for-Relative-Effectiveness-Assessment-Health-related-quality-of-life-and-utility-measures_Amended-JA1-Guideline_Final-Nov-2015.pdf		changes
ISPOR – The Professional Society for Health Economics and Outcomes Research	15-17	5.3 Interpretability of scales (Lines 459-521, inclusive of box)	It would be helpful to clarify the current language regarding MID/MCID to more clearly delineate between group-level and patient-level. In addition, the recent FDA PFDD Guidance 3 workshop introduced the term "meaningful within-patient change (MWPC)," which may be relevant to note. There are also references on responder analyses minimal differences that may be worth citing, such as the recent FDA guidances, Revicki et al (2008), King (2011) and Coon & Cappelleri (2016).		The authors are aware of the many, old, but still ongoing debates about the terminology within the literature about interpretability (meaningful versus minimal, difference versus change, clinical or not clinical, between group versus within group…). Minimal Perceived Change has also been recently proposed, with the support of a conceptual model which for the first time defines this concept as a statistical parameter with a population-level definition. While all these debates are important, we have nevertheless retained the term MID as the guideline has a practical purpose first and MID is still the most prevalent term in the literature. We will nevertheless add a clarification that we use the term in the guideline in the context of estimating a RD threshold that helps to interpret within patient-change.
ISPOR – The Professional Society for Health	15-17	5.3 Interpretability of scales (Lines 459-521, inclusive of box)	There is a recent paper in Value in Health which is useful for this section. To help interpret composite endpoints which use a responder definition (Lines 447-484), it shows how you can link the response criteria to changes in		Thank you for the reference.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Economics and Outcomes Research			health utility values used by health technology assessment (HTA) bodies. Producing this evidence will help HTA agencies interpret whether the endpoint corresponds with health gain valued according to their preferred instrument (which varies between member state jurisdictions). Paper details here: https://doi.org/10.1016/j.jval.2022.07.001		
ISPOR – The Professional Society for Health Economics and Outcomes Research	15-17	5.3 Interpretability of scales (Lines 459-521, inclusive of box)	Several important references on interpretation are missing and merit citation: Cappelleri JC, Bushmakin AG. 2014. Interpretation of patient-reported outcomes. Statistical Methods in Medical Research 23:460-483. Coon CD, Cappelleri JC. 2016. Interpreting change in scores on patient-reported outcome instruments. Therapeutic Innovation & Regulatory Science 50:22-29. Coon CD, Cook KF. 2018. Moving from clinical significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. Quality of Life Research 27:33-40. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. 2007. Understanding the minimum clinically important difference: a review of concepts and methods. The Spine Journal7:541-546. King MT. 2011. A point of minimal important difference (MID): a critique of terminology and methods. Expert Review of Pharmacoeconomics and Outcomes Research11:171-184. McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. 2011. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert Reviews of		These references are known by the authors and are of course sound literature. Nonetheless, the purpose of the guideline is first to provide practical guidance to assessors for producing JCA report and not to be a systematic review of the topics that are briefly summarized within the guideline.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			Pharmacoeconomics & Outcomes Research 11:163–169. Revicki D, Hays RD, Cella D, Sloan J. 2008. Recommended methods for determining responsiveness and minimally differences for patient-reported outcomes. Journal of Clinical Epidemiology 61:102-109. Patient-reported outcome measures (PROMs) of an underlying continuous nature should be primarily analyzed as continuous outcomes to detect treatment effect, and responder analyses should be reserved as secondary analyses for enhancing interpretability and for regulatory purposes of PROMs. Here are supportive references: Collister D, Bangiwala S, Walsh M, Mian R, Lee SF, Furukawa TA, Guyatt G. Patient reported outcome measures in clinical trials should be initially analyzed as continuous outcomes for statistical significance and responder analyses should be reserved as secondary analyses. J Clin Epidemiol 2021; 134:95-102. Cappelleri JC. Further reduction in statistical power for responder analysis of patient-reported outcomes with measurement error. Journal of Clinical Epidemiology. 2021; 140:200-201.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	15-17	5.3 Interpretability of scales (Lines 459-521, inclusive of box)	Patient-reported outcome measures (PROMs) of an underlying continuous nature should be primarily analyzed as continuous outcomes to detect treatment effect, and responder analyses should be reserved as secondary analyses for enhancing interpretability and for regulatory purposes of PROMs. Supportive References: Collister D, Bangiwala S, Walsh M, Mian R, Lee SF,		We agree. It will be clarified in the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			Furukawa TA, Guyatt G. Patient reported outcome measures in clinical trials should be initially analyzed as continuous outcomes for statistical significance and responder analyses should be reserved as secondary analyses. J Clin Epidemiol 2021; 134:95-102. Cappelleri JC. Further reduction in statistical power for responder analysis of patient-reported outcomes with measurement error. Journal of Clinical Epidemiology. 2021; 140:200-201.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	15-17	5.3 Interpretability of scales (Lines 459-521, inclusive of box)	Suggest adding the term “Meaningful Within-Patient Change (MWPC)” in the guideline. MWPC has been illustrated in the FDA Guidance PFDD Public Workshop Guidance 3 Discussion Document (fda.gov). The current statements do not distinguish the patient-level change (e.g., MWPC) and the group- level difference (e.g., mean difference between treatment groups). It is suggested to clarify the terms in patient-level or group-level when “MID” and “MCID” are used. It would better to align with the recommendations in the FDA Guidance PFDD Public Workshop Guidance 3Discussion Document (fda.gov).		Already addressed issue. See answer above on the debates on terminology about interpretability.
ISPOR – The Professional Society for Health Economics and Outcomes Research	15-17	5.3 Interpretability of scales (Lines 459-521, inclusive of box)	FDA recommends the use of anchor-based methods supplemented with both empirical cumulative distribution function (eCDF) and probability density function (PDF). Please consider adding reference to the recent set of FDA guidances --https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical.		Use of CDFs are highly encouraged within the guideline.
ISPOR – The Professional	15-17	5.3 Interpretability of scales (Lines 459-	Please comment on whether 2 studies (clinical trials) are required to establish MCID and measure the response		The guideline asks for factual reporting of elements of methods and results that allow

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Society for Health Economics and Outcomes Research		521, inclusive of box)	analysis of achieving MICD by treatment		MS to draw their own conclusion. The main recommendation is to ask HTD to provide the elements that allows to understand how a RD was obtained, if any is provided in the evidence submitted. Nonetheless, the appraisal of the soundness of this RD is left at the discretion of the MS, and the purpose of the guideline is not to endorse how MIDs must be estimated.
ISPOR – The Professional Society for Health Economics and Outcomes Research	15-17	5.3 Interpretability of scales (Lines 459-521, inclusive of box)	Overall, a recommendation about adequate determination of responders would be useful. Of note, validated, established response thresholds are useful and valuable to enhance the transferability of risk/benefit assessments based on PRO measures to assess relative effectiveness for health technology assessment and to ensure consistent interpretation of PRO effects. A singular threshold of x%-change of the continuous scale range for all instruments is incongruent with previously defined and scientifically established thresholds and is not well-suited for universal implementation. [Reference: Schlichting et al, Is IQWiG's 15% Threshold Universally Applicable in Assessing the Clinical Relevance of Patient-Reported Outcomes Changes? An ISPOR Special Interest Group Report, Value in Health, Volume 25, Issue 9, 2022, Pages 1463-1468, ISSN 1098-3015, https://doi.org/10.1016/j.jval.2022.07.010. Please add some language as follows. “In general, validated and established response thresholds should be		We will consider if clarifications are needed for the next version of the draft, but we do think the draft already conveys the idea that anchor-based methods are the only ones that take into account the patient’s perspective, and the guideline does not strongly support a unique threshold and outline the shortcomings of distribution-based RDs.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			considered when defining relevant individual response thresholds in terms of MID. Anchor based methods to determine MIDs utilizing patient-reported anchors are preferred. In absence of patient-reported anchors clinician reported anchors should be considered."		
ISPOR – The Professional Society for Health Economics and Outcomes Research	5-6	1.2 Relevant articles in Regulation (EU) 2021/2282 (Lines 117-123)	Providing links to these articles would be helpful.		It will be considered for the next version of the draft.
ISPOR – The Professional Society for Health Economics and Outcomes Research	6-7	2.1 Definitions (Lines 124-177)	It would be helpful to mention the term clinical outcome assessment (COA) and align with existing and commonly used definitions of specific COAs such as patient-reported outcomes (PROs), clinician-reported outcomes (ClinROs), performance outcomes (PerfOs), proxy-reported outcomes (ProxROs) and others. In addition, it is worth nothing that there have been professional society task forces and documents that have discussed COAs such as ClinROs, PerfOs, and ProxROs. ISPOR Task Forces have produced reports to help standardise ClinROs and PerfOs and ISOQOL has a Task Force on proxy reporting/ProxROs with a similar objective. Furthermore, it would be worthwhile to note existing regulatory definitions around COA terms such as those used by FDA and EMA. Although it is certainly relevant to note the growth and importance of digital data, it may be worth making a distinction between digital data and PerfOs; as currently		Clarifications will be proposed for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			written, the paragraph could be interpreted as suggesting an inherent link between the two. In addition, the proposed definition of proxies is relatively non-standard, and it may be beneficial to cite regulatory or other existing definitions of proxies. Furthermore, although carers/caregivers can provide proxy reports, in some situations the carer/caregiver experience on its own may be relevant, and thus caregiver outcomes could be mentioned. Distinguishing between proxy-reported outcomes and observer-reported outcomes would also be helpful.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	10-11	3.3 Surrogate outcomes (Lines 285-336)	Statement in the box, bullet point 1 "If evidence for a patient-centred outcome is likely to be available, then this should be requested during the scoping process instead of surrogate outcomes such as morbidity, overall mortality and HRQoL" is inconsistent with ll.224-225 which names mortality as a patient-centered outcome, as well as with the usual treatment of morbidity, mortality and HRQoL as ultimate outcomes, depending on the condition.		We will consider if this precision is necessary for the next version of the draft.
ISPOR – The Professional Society for Health Economics and Outcomes Research	14-15	5.2 Validity and reliability of scales (Lines 406-458, inclusive of box)	It would be beneficial to mention other considerations for PROMs, including fitness for context of use and the potential relevance of conceptual frameworks to guide the selection of appropriate concepts. In addition, the use of the term 'content validity' could be helpful as well, given its importance in overall PROM validity. There are several well-known and well-cited books that may be useful to cite to guide readers to helpful references, specifically Cappelleri et al (2016) and Streiner et al (2015).		We will consider clarifications for the next version of the draft.
ISPOR – The Professional	14-15	5.2 Validity and reliability of scales	7th bullet – table: For PROMS this description should also include evidence that the instrument is fit for the		We will consider it if this is necessary for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Society for Health Economics and Outcomes Research		(Lines 406-458, inclusive of box)	context of use.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	14-15	5.2 Validity and reliability of scales (Lines 406-458, inclusive of box)	Table on p. 15, last bullet: Regarding the following sentence: “References, as provided by the HTD, allowing the access to the specific (clinical) studies assessing the measurement properties (and measurement model) of the instruments that are used.” Please clarify what kind of references should be provided (published, data on file).		We will consider if this clarification is necessary for the next version of the draft.
ISPOR – The Professional Society for Health Economics and Outcomes Research	14-15	5.2 Validity and reliability of scales (Lines 406-458, inclusive of box)	"Content validity" is not discussed. Psychometrics (eg, l. 432) is important to ensure the statistical properties of a questionnaire, but the first step is to be sure 1/ that the content of the questionnaire is really measuring the concepts it intends to capture and 2/ that the concepts are relevant for the patients included in the study		We agree that content validity is an important component of validity, but the purpose of the guideline is not to be a textbook on clinimetrics/psychometrics. Its primary purpose are the recommendations for assessors and co-assessors.
ISPOR – The Professional Society for Health Economics and Outcomes Research	14-15	5.2 Validity and reliability of scales (Lines 406-458, inclusive of box)	There are a few other noteworthy, well-cited books that deserve to be referenced: Cappelleri JC, Zou KH, Bushmakin AG, Alvir JMJ, Alemayehu D, Symonds T. 2013. Patient-Reported Outcomes: Measurement, Implementation and Interpretation. Boca Raton, Florida: Chapman &Hall/CRC Press. Fayers PM, Machin D. 2016. Quality of Life: The Assessment, Analysis and Reporting of Patient-Reported Outcomes. Third edition. Chichester, United Kingdom: John Wiley & Sons Ltd. Streiner DL, Norman GR, Cairney J. 2015. Health		Fayers et al. is already cited within the guideline. The guideline is not meant to be a systematic review, but primary to provide practical guidance for assessors. Regarding random errors, the sentence only refers to reliability.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			Measurement Scales: A Practical Guide to Their Development and Use. Fifth edition. New York, NY: Oxford University Press. Line 422: There can be systematic error as well as random error, so I might drop (i.e., random error) at the end of the sentence.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	14-15	5.2 Validity and reliability of scales (Lines 406-458, inclusive of box)	In empirical studies, the reliability of the measure can place limits on the empirical validity results, meaning that in some cases establishment of reliability may need to precede some aspects of the validity work.		We agree with this comment, but intricacies or hierarchy between validity and reliability are technical considerations that are beyond the purpose of this practical guideline.
ISPOR – The Professional Society for Health Economics and Outcomes Research	8-9	3.1 Definition of patient-centred outcomes (Lines 210-250, inclusive of box)	It would be helpful to differentiate patient-centered care with physician centered care using examples. In addition, it would be helpful to suggest the 'essential' vs 'desirable' outcomes needed for JCA to HTDs. While OS is relevant, it may be infeasible for HTDs to measure OS in early-stage cancers. It is equally important not to ignore patient-relevant surrogate outcomes. Recommending the use of validated tools only might pose challenges in certain therapeutic areas where such validated measures are not available. It also limits the innovative use of endpoints or measures from RWD sources, even if only as exploratory endpoints.		Thank you for your comments.
ISPOR – The Professional Society for Health	8-9	3.1 Definition of patient-centred outcomes (Lines 210-250, inclusive	What is the difference between the "patient-centered outcomes" presented here and the "patient-relevant outcomes" presented in previous guidelines? Also, can "patient-centered outcomes" be considered to include		Patient-centred outcomes do indeed include PRO in its definition (see definition in the guideline). Patient-centred and patient relevant can be considered similar

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Economics and Outcomes Research		of box)	PROs?		concepts but we there was a need to choose one term and patient-centred was retained because its abbreviation do not lead to “PRO” while it can be the case with patient-relevant outcome which could lead to confusion with patient-reported outcome.
ISPOR – The Professional Society for Health Economics and Outcomes Research	7-8	178-210	The wording for the scoping process on the outcomes should be accompanied by justification/rationale, in keeping with good statistical practice, to enable HTD to better understand the request. These should also be aligned to the study design, otherwise it becomes an immediate disadvantage to the HTD for not being able to provide them.		Thank you for your comments. We will consider it.
ISPOR – The Professional Society for Health Economics and Outcomes Research	7-8	178-208	It would be good to consider and clearly state RWE outcomes as well as the outcomes from RCTs in these three paragraphs.		Thank you for your comment
ISPOR – The Professional Society for Health Economics and Outcomes Research	7-8	2.2 General considerations (Lines 178-208)	In this part or somewhere, is there any definition of difference between preference-based measures (PBM) and non-PBM (including COS) or utility?		We are not sure to understand what the issue at stake is precisely.
ISPOR – The Professional	12-13	4.3 Information to be reported for	It may be useful for the document to introduce the notion of self-reported AE (using e.g. the PRO-CTCAE).		Already addressed issue

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Society for Health Economics and Outcomes Research		safety outcomes (Lines 355-380, inclusive of box)			
ISPOR – The Professional Society for Health Economics and Outcomes Research	12-13	4.3 Information to be reported for safety outcomes (Lines 355-380, inclusive of box)	There is a risk of bias in the causality determination in blinded studies, because blinding is rarely perfect (e.g., difference is AE or clinical response may give the investigator an idea of whether a patient is on control or experimental treatment); the bias is just greater in unblinded studies. Since the conclusion applies to both blinded and unblinded studies, the second sentence does not help. Consider the following alternative statement: "Causality (attributability) between a heath technology and an AE could be described by many terms and scales. However, there is always uncertainty and risk of bias in of the determination of "causality", so all safety outcomes must always be reported, irrespective of causality designation."		We do not think the proposed sentence will improve the readability of the guideline.
ISPOR – The Professional Society for Health Economics and Outcomes Research	13-14	5.1 Definitions and general considerations (Lines 381-405, inclusive of box)	It may be helpful to rephrase the statement regarding PROs being "less objective" to avoid suggesting that PROs are less valuable compared to technological or performance measures. Some would say the major value of including PRO's is to capture the perspective of the patient in a subjective sense. In addition, ObsROs and ProxROs may be worth mentioning in this section, as in some cases patients may be unable to complete PROs.		"Objective" and "subjective" are not used here in their colloquial meanings ("true" or "valid" as opposed to "invalid") but refer to the necessity or not of involving a subjective judgment of a person in the process of producing the measure. The purpose of the paragraph was not to discuss measurement properties, nor to discard PROMs, but to emphasize first that there exist measures that do not involve such subjective assessment (e.g., the objective height of a person as measured in cms) vs ones that do (e.g., does a

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
					person feel he/she small or tall?) and secondly that if one wants to have insights on patients' preferences, values, judgments, they need to use a measure which allows the access to these internal standards. Clarifications will be made in the next version of the draft.
ISPOR – The Professional Society for Health Economics and Outcomes Research	13-14	5.1 Definitions and general considerations (Lines 381-405, inclusive of box)	1st bullet – table at top of page 13: While "medical technology" can be considered to cover passive measures such as digital health technologies, please consider including passive technologies in any more detailed discussion about medical technology as a data source.		The paragraph on digital outcomes will be reworded for the next version of the draft.
ISPOR – The Professional Society for Health Economics and Outcomes Research	13-14	5.1 Definitions and general considerations (Lines 381-405, inclusive of box)	Table at top of p.13 - in addition to the "main source of information" (prior validation evidence package), there should be a conceptual framework to show why the concepts (symptoms/ impact/HRQoL) are identified to be covered by the PROMs selected in the trial.		We will consider if this addition is necessary for the next version of the draft.
ISPOR – The Professional Society for Health Economics and Outcomes Research	15-16	471-476	Interpretability of scales may vary in case categorical scales are transformed into continuous scales or vice versa. Whether a continuous or categorical scale is used to determine the endpoint of interest strongly depends on the underlying objective. One might be interested to explore the rate of patients who improved / maintained / worsened their symptoms compared to baseline (patient level objective), or the average change compared to		This will be considered for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			baseline (within group perspective). It is difficult to argue which of the associated analyses complements the other as they address different questions. The risk of data dredging can be avoided by specifying key outcomes of interest in advance of the analysis which is typically done in the analysis plans. It is recommended to align key outcomes with the agencies in joint scientific advice meetings. The inflation of type I error rate might not be an exclusive problem here but also for subgroup analyses. Please consider changing the language as follows: "this expression of treatment effectiveness can enhance interpretability. Analysis on the categorical scale could complement the analysis on the continuous scale and vice versa. In addition, to avoid the risk of data dredging and, one measure of treatment effect should be pre-specified in the protocol and statistical analysis plan as a primary analysis (see the EUnetHTA 21 practical guideline "Applicability of evidence: practical guideline on multiplicity, subgroup, sensitivity and post-hoc analyses").		
ISPOR – The Professional Society for Health Economics and Outcomes Research	10-11	287-291	A surrogate outcome is a measurement that is not the ultimate/relevant outcome of the disease (biomarker, measure of a function) that has proven to be correlated with the ultimate endpoint of the disease. It should be stated explicitly that any biomarker or intermediate outcome used as a surrogate outcome must have demonstrated such a correlation – not all biomarkers or intermediate outcomes necessarily do so. Further, the biological or other plausibility of the causal pathway from surrogate marker to ultimate endpoint should be stated; otherwise, a statistically significant association between		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			the surrogate outcome and the patient-centered outcome might only indicate confounding. Surrogates may also be used to address issues of confounding in longer term outcomes. E.g., PFS as treatment effect in oncology trials to avoid issues with OS from crossover/subsequent therapies etc.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	10-11	304-308	This paragraph may be strengthened and better justified by adding "due to the smaller sample sizes and shorter duration often associated with trials employing surrogate markers" to the first sentence. These factors create some uncertainty about whether adverse event risk has been fully captured, which, along with potential greater uncertainty about clinical benefit, are the main considerations affecting risk-benefit evaluation. However, long treatment duration is not always relevant. Progression free survival is used as a proxy for overall survival in oncology trials and the requirement of a benign safety profile is less applicable to oncology treatments.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	8-9	223-224	To understand the "contrast" meant here, it would be helpful to define physician-centred care. Do you mean clinician-reported outcomes, or physician services as typically captured in quality of care measures, or something else? Mortality is clearly important to patients, but is generally clinician-reported. Thus understanding what is meant by physician-centered care will help clarify the contrast with, and meaning of, patient-centered outcomes.		
ISPOR – The Professional Society for	7-8	2.2 Summary	It may be beneficial to extend the recommendation regarding request formulation to mention international standards such as SISAQOL (standardisation of PRO		We will consider if clarifications are necessary for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Health Economics and Outcomes Research			analyses in cancer trials). Furthermore, rephrasing the recommendation to suggest that good clinical and statistical practice be incorporated would help to remind readers of the importance of these issues. Additionally, it would be worth mentioning the importance of pre-specifying statistical methods as part of these general recommendations.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	4	73	Missing acronyms that could be interesting to add and to mention in the text: - PRO-CTCAE, as FDA and EMA acknowledge the value of self-report of symptomatic AE by patients to complement the physician report of AE - apart of the PGRC, another global assessment is the PGIS: patient global impression of severity; COA – clinical outcome assessment. EQ-5D is missing		These will be considered for the next version of the draft.
ISPOR – The Professional Society for Health Economics and Outcomes Research	5	1.1 Problem statement, scope and objectives (Lines 75-116)	It isn't clear if the guidance is proposing a set of harmonized recommendations that all member states need to follow or if the guidance is allowing member states to select certain recommendations that are more applicable and relevant to them. It is critical to strike a balance in proposing harmonized recommendations and accommodating the varying requirements of each of the member states at national level. For JCA, aspects pertaining to measurement scales / instruments for assessing outcomes including PROs need to be harmonized across member states.		We do think it is clear that the two types of recommendations are made: 1/ guidance for helping MS to express their needs regarding outcomes in the context of the HTAR (no ranking and no rationale for requests) 2/ what needs to be reported within a JCA.
Pooyeh Graili	5	77-99/1	Unclear		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Quality HTA					
ISPOR – The Professional Society for Health Economics and Outcomes Research	5	100-105	This guideline can also be useful for the submitters.		
Pooyeh Graili Quality HTA	5	100-105/1	This guideline can also be useful for the submitters.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	5	85-88	The judgement of what constitutes a clinically relevant outcome, responder definition, etc. should be harmonized across member states (MS) based on a scientific rationale. Please consider changing the sentence as follows: “While MS are required to give due consideration to the JCA reports published (Article 13 (1)), the rating of the additional benefit of a treatment may differ at a national level which is based on the clinical relevance of the measure of relative effectiveness.		The HTAR does not call for such harmnonization.
ISPOR – The Professional Society for Health Economics and Outcomes Research	5	96-99	The guideline says that “the conclusions that MS can draw regarding the clinical added value of a treatment can be impacted by factors such as appraisal of the validity and reliability of the measurement scales of instruments or of the relevance of intermediate or surrogate outcomes.” The assessment of the validity and reliability of the measurement scales of instruments should be based on scientific standards and should be harmonized across member states.		Based on scientific rationale or evidence-based do not imply that decision-making is devoid of any kind of appraisal or debate. Validity, reliability and interpretability of either outcome measurement instruments or surrogate outcomes are not on/off concepts. They come in degrees. This is why the guideline aims for a factual reporting of the methods and results elements regarding these aspects, but MS can appraise them the way they see fit.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
ISPOR – The Professional Society for Health Economics and Outcomes Research	5	91	Does this refer to rank ordering of endpoints? If so - this is contradictory to what regulatory agencies usually require (rank ordering and prespecifying as a prerequisite to granting labeling language around those endpoints). Also in reference to ranking on health outcomes, including weighting and relevance of primary vs. secondary vs. exploratory endpoints: Often, payer-related endpoints are secondary or exploratory given the trial is not powered to show differences. How will these endpoints be considered in the evaluations? In addition to the JCA assessments, will member states require or ask for data for patients in their countries? What if there are no patients in the trials for that country?		It refers to the fact that outcomes that MS will request during the assessment scope will not be ranked.
ISPOR – The Professional Society for Health Economics and Outcomes Research	6	161-166	It might be better to distinguish “observer-reported outcome” and “proxy-reported outcome” here. “...reported by an observer with shared experience. An example would be a caregiver if the patient is unable to answer the items. These cases are referred to as PROs answered by "proxies” ….”		This will be clarified.
ISPOR – The Professional Society for Health Economics and Outcomes Research	6	147-151	Suggest adding the acronym here (ClinRO). The term "clinically reported outcomes” is inconsistent with the sentence before it (Clinician-reported outcomes), which is the usual term for these measures. We encourage use of the Clinician-reported outcomes, or ClinRO, term, and suggest these measures be referred to in a consistent manner.		This will be clarified.
ISPOR – The Professional	6	140-143	It seems “absolute effect” is one type of difference measure. If so, we may reword it as below.		This will be clarified.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Society for Health Economics and Outcomes Research			"… effect measures are either difference measures (e.g., absolute effect, mean difference in change, risk difference) or ratio measures (e.g., risk ratio, odds ratio, hazard ratio). However, other statistics can be used to express other aspects of a treatment effect such as the absolute effect or a within-group change."		
ISPOR – The Professional Society for Health Economics and Outcomes Research	6	126-127	Suggestion for lines 126 and 127: Health outcome is the impact that a specific health intervention, technology, policy, or program has on a person, group, or population. This endpoint can be any changes in morbidity (efficacy, effectiveness, and safety) and/or mortality. Safety and effectiveness impact on HRQOL and QOL.		We think the current introduction is fine.
ISPOR – The Professional Society for Health Economics and Outcomes Research	6	138-139	138 and 139: instead of existing 'health technologies', we can say 'health interventions' because the comparator can be a surgical method or a public health intervention, not a health technology.		This is a direct quote of the HTAR.
Pooyeh Graili Quality HTA	6	126-127/2	Health outcome is the impact that a specific health intervention, technology, policy, or program has on a person, group, or population. This endpoint can be any changes in morbidity (efficacy, effectiveness, and safety) and/or mortality. Safety and effectiveness impact on HRQOL and QOL.		We think the current introduction is fine.
Pooyeh Graili Quality HTA	6	138-139/2	Instead of existing 'health technologies', we can say 'health interventions' because the comparator		Duplicated comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			can be a surgical method or a public health intervention, not a health technology.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	6	152	Technology assessed outcome measurements could be made through digital solutions, diagnostic tools (cont glucose measurements), AI and algorithms. To some extent they are mentioned in a section further into the document, however, they could also be mentioned under technology assessed outcomes		It will be clarified.
ISPOR – The Professional Society for Health Economics and Outcomes Research	6	164	Can there be recognition of Caregiver reported outcomes, as a standalone measure and resulting from the intent to measure the caregiver perspective (i.e separate from their role as a proxy?). This will allow assessment of the broader impact of a condition beyond individual patients' experiences		It will be clarified.
ISPOR – The Professional Society for Health Economics and Outcomes Research	7	192-201	"Timing of outcome assessment" should be justified in the context of the kind of trial and the treatment and follow Good Clinical and Statistical Practice. For example, it could be inappropriate to ask for an OS delta in neoadjuvant treatment at 12 months because other endpoints are more appropriate.		We agree.
ISPOR – The Professional Society for Health Economics and Outcomes Research	7	195-198	The guidance addressed concerns if the follow-up was considered not sufficiently long in the clinical study submitted as evidence and recommended to formulate a request "[Outcome of interest] measured preferably at [insert timing of assessment]". Preferences may vary between MS. One MS might be interested in ""rate of major adverse cardiovascular events 2 years after inclusions" another MS is interested in 12		We will consider if a change is needed for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			months or 18 months, 3 years after inclusion. On the other hand, the data collection of a study usually covers the clinically most important periods, where changes in symptoms or side effects are expected. In addition, data collection period and frequency of assessments also consider the factors that would compromise a scientifically sound evaluation e.g. when the rate of missingness is considered too high to draw reasonable conclusions and also acceptable level of patient burden. In essence the request for a timing of assessment should not only be harmonized across member states, it should also adequately reflect the disease setting and clinical context. So, a joint scientific advice meeting (including REG and HTA) is recommended to clarify the needs so that the study can be designed accordingly. Second, as marketing authorization could be based on a positive benefit-risk evaluation using interim results of a study, only sparse data might be available to provide a reasonable precise estimate e.g., for the rate of major adverse cardiovascular event 2 years after inclusion. Modelling approaches may be informative to estimate a specific outcome at a certain time point accordingly. Such methods could be useful to decrease uncertainty and thus should be taken into account for the added benefit assessment. In summary, please consider changing the wording from: "A general recommendation could also be to formulate a request as such: "[Outcome of interest] measured preferably at [insert timing of assessment]". To "A general recommendation could be to align key time points of interest prior to study initiation in a joint scientific advice meeting, or to pre-specify statistical methods to		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			estimate the effect at certain time points also using modeling approaches. “		
ISPOR – The Professional Society for Health Economics and Outcomes Research	7	183-184	The first sentence of the paragraph is unclear, mainly with regard to whether it refers to the MS or the HTD. Also, sometimes results are not best 'obtained‘ from broader concepts, but rather are better 'obtained' from specific items or subscales of HRQoL.		The sentence does not refer to generic versus specific HRQoL. It will be clarified for the next version of the draft.
ISPOR – The Professional Society for Health Economics and Outcomes Research	7	191-192	General considerations refer to the definition of an outcome, in particular specifying an appropriate PRO endpoint. The guideline says, “To alleviate this issue, a general recommendation could be to formulate a request as such: “[Outcome of interest] measured preferably as [insert measure]”.” The recommendation should be extended, and reference given to the SISAQOL-IMI2 initiative aiming for setting International Standards in Analysing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials. Similarly, to the reference to COMET for COS funded also by IMI2. Thus, please consider adding “To alleviate this issue, a general recommendation could be to formulate a request as such: “[Outcome of interest] measured preferably as [insert measure]” and to consider international standards in analyzing PROs and Quality of Life Endpoints in cancer clinical trials as provided by SISAQOL-IMI.”		HTAR is not limited to cancer trials. We will consider if endorsing such a guidance is necessary for this guideline.
ISPOR – The Professional Society for	7	195-196	There should be recognition that with PROs there are limitations in how long they can be assessed within a study. Also, general flexibility for HTDs in designing the		The guideline is not intended to describe every difficulty that can occurred during clinical developments regarding the

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Health Economics and Outcomes Research			schedule of assessment in their studies would be appreciated, as treatments may have different onset of action. The schedule of assessments should reflect that difference.		collection of outcomes.
Pooyeh Graili Quality HTA	7	178-208/2	The subtitle could be something more relevant to the content rather than 'general considerations'. It is good to consider and clearly state RWE outcomes as well as the outcomes from RCTs in these three paragraphs.		The guideline proposes recommendations independently of the source of evidence.
	7	176	In addition to DAS, the Mayo Score is another hybrid measure that can be added. The Mayo Score was developed as a composite disease activity index for use in clinical trials. The original description of the Mayo Score included an assessment of 2 patient-reported outcomes [PROs; stool frequency (SF) and rectal bleeding (RB)], the endoscopic appearance of the mucosa (endoscopic score, ES), and a Physician's Global Assessment (PGA), each of which were scored on a scale from 0 to 3, giving a maximum total score of 12. Reference: Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. N Engl J Med.1987; 317:1625–1629.		DAS was only introduced as an example. The purpose of the paragraph is not to endorse a list of validated scores.
ISPOR – The Professional Society for Health Economics and Outcomes Research	7	190	Does "assessed differently" mean using a different instrument? Please be more specific.		This will be clarified.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Pooyeh Graili Quality HTA	8	Bullet one/Summary tabel	Impact instead of 'effectiveness' because sometimes the outcome of an intervention is fewer side effects but not effectiveness. Considering the defined outcomes in this guideline are based on the evidence from RCT, we also need to elaborate on the outcomes if from RWD sources.		While we understand the comment, we will keep the term effectiveness as it is the main purpose of the HTAR. The guideline provides recommendations independently of the sources of evidence. Further guidance on RWD is provided in the D4.6 guideline.
ISPOR – The Professional Society for Health Economics and Outcomes Research	8	212-220	We acknowledge that outcomes supporting the benefit-risk assessment might be "less suitable for the needs of JCA". However, the healthcare and treatment decisions strongly depend on the prescribing information (PI). Hence all outcomes described in the PI should be considered for JCA purposes and used as a common "core" outcome set among all member states. Please consider changing the sentence to "Some outcomes may be fully acceptable as support for the risk/benefit ratio assessment of a certain therapy but are less suitable for the needs of JCA. However, outcomes described in the prescription information should be considered for JCA purposes." "The acceptability of an outcome is subject to MS interpretation of their relevance within their national process for decision making and thus may differ between MS. At least outcomes described in the prescription information should be considered as a core set of outcomes. "		
ISPOR – The Professional Society for	8	215-216	How does this apply to an acute condition, e.g., infection?		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Health Economics and Outcomes Research					
ISPOR – The Professional Society for Health Economics and Outcomes Research	8	2 Summary Table (Line 209)	With reference to the following sentence “Effect measures should not be specified by MS. The HTD is responsible for presenting results using appropriate effect measures in accordance with good clinical and statistical practice”. Please clarify the criteria/guidelines that the HTD should consider when selecting effect measures.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	8	2 Summary Table (Line 209)	Bullet 1: Should “or safety” be added to the end of this sentence?		
ISPOR – The Professional Society for Health Economics and Outcomes Research	8	2 Summary Table (Line 209)	General comment to the requirement for member states in the assessment scoping process – please include a request for the rationale underlying their requests for specific outcomes and measures.		
ISPOR – The Professional Society for Health Economics	8	2 Summary Table (Line 209)	The last bullet is not worded clearly. We suggest rewording as follows: “An accurate definition of any reported outcome is required and would include a description of the concept, source of information, the measure of the outcome, timing, and effect measure.”		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
and Outcomes Research					
ISPOR – The Professional Society for Health Economics and Outcomes Research	8	224	Anchoring on the emphasis on patient-centered outcomes as those which matter to patients, one should try to bring in the relevance and importance of surrogate outcomes (not withstanding what has been said above in line 213-215 about surrogate outcomes) because of the importance that patients might ascribe to those outcomes in early-stage disease (i.e., oncology).		
ISPOR – The Professional Society for Health Economics and Outcomes Research	9	236-242	Marketing authorization of a drug could be based on a positive benefit-risk evaluation using interim results of a study based on dual endpoints one of which demonstrated superiority (e.g. PFS in a cancer study) and the other (overall survival OS) show at least a clear trend. A final OS outcome is rarely available as studies are still ongoing at the time of JCA. Additional endpoints such as PFS2 (time to to second objective disease progression, or death from any cause, whichever first) and time to next anti-cancer treatment could be considered as intermediate endpoints to substantiate the trend observed for an interim OS outcome. Interim analyses may provide only sparse data to provide a reasonable precise estimate for long term survival (e.g., 5-years survival rate). intermediate endpoints and modelling approaches could be useful to decrease uncertainty of long-term outcomes and thus should be taken into account to assess the added value of a health technology. Another aspect concerns maintaining the study integrity of a trial. For instance, if interim results of a randomized controlled double blind clinical study qualify for marketing		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			authorization, independent data monitoring committees may suggest keeping the double-blind nature of the trial to provide an unbiased OS estimate. HTDs may have to decide whether to risk losing a full approval or a positive benefit assessment by MS when unblinding based on strong intermediate outcomes. Early access to an efficacious drug could be compromised by conflicting requirements and thus decisions might not be considered patient centered. Please consider changing the following sentence as follows: If it is not feasible to measure a final outcome, then intermediate or surrogate outcomes may be acceptable. Acceptability should consider if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome [14], study integrity considerations e.g. when studies need to be unblinded to obtain the final outcome and pre-specified statistical approaches to estimate the long-term using modelling approaches. Scientific advise meetings are recommended to clarify opportunities.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	9	266-272	The recent FDA draft guidance, “Core Patient-Reported Outcomes in Cancer Clinical Trials,” would be a most helpful complement to this paragraph: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/core-patient-reported-outcomes-cancer-clinical-trials.		Thank you for your comment.
ISPOR – The Professional	9	239-243	Please develop a method/matrix table to clearly specify which outcomes will be considered as “essential” for HTA		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Society for Health Economics and Outcomes Research			evaluation, depending on the specific disease (e.g. metastatic vs. early-stage cancers). This would help HTDs develop clinical trial protocols to collect outcomes that are meaningful for HTA evaluations.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	9	277-280	1. This is the only mention of generic multiattribute utility instruments in the document; more context about their use as endpoints would be helpful. 2. Given their importance in creating QALYs for CEA – even if not JCA – we would encourage their inclusion as complementary endpoints in trials *whenever possible*.		Thank you for your comment.
ISPOR – The Professional Society for Health Economics and Outcomes Research	9	252-254	It would be most helpful to provide a method/matrix table to clearly specify which outcomes will be considered as "essential" for HTA evaluation, depending on the specific disease (e.g. Metastatic vs. early-stage cancers). This would help HTDs appropriate develop clinical trial protocols to collect outcomes that are meaningful for HTA evaluations.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	9	236	Please provide a more inclusive definition of long-term outcomes, since not all conditions are chronic or fatal.		
ISPOR – The Professional Society for	9	242	Alternatively (to the suggestion of strong correlation), given the acknowledgment around patient-centered care in the paragraph starting line 223, surrogate outcomes		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Health Economics and Outcomes Research			need to be considered acceptable if patients find them meaningful.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	9	248	Would avoid mention of 'validated' in this context in which clinical scales are referred to. A large proportion of such clinical measures aren't validated relative to the rigor of similar efforts expected of PROs. Alternatives: validated tools...or those developed by professional clinical associations and guidelines developers.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	9	250	In the “Points of attention for the assessment of scoping process” box: Given the acknowledgment around patient-centered care in the paragraph starting line 223, surrogate outcomes need to be considered acceptable if patients find them meaningful. “If it is not feasible to measure final outcomes, then intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome.” Suggestion: If it is not feasible to measure final outcomes, then intermediate or surrogate outcomes may be acceptable if there is evidence of a strong association or correlation of effects on the surrogate or intermediate outcome with the effect on the final outcome OR evidence that patients find the intermediate and/or surrogate outcomes to be meaningful.		
ISPOR – The	9	250 (box)	To reiterate, long-term or final outcomes do not seem defined widely enough to cover all conditions.		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Professional Society for Health Economics and Outcomes Research					
ISPOR – The Professional Society for Health Economics and Outcomes Research	9	259	If “multimorbidity” is meant as a specific situation and not simply as a general example, it would be helpful define a "multi-morbidity" condition in terms of the combination, severity, and number of conditions.		Thank you for your comment. This will be clarified in the next version of the guideline.
Pooyeh Graili Quality HTA	10	297/I	irreversible morbid events is misleading because the ultimate morbid outcome in some disease is reversible.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	10	308-309	Points of attention for the assessment scoping process. Often so-called surrogate endpoints are not intended to replace a patient-centered outcome but are important to supplement or complement the assessment of the patient-centered outcome. They could add value as they could limit gaps related to the certainty of the outcome. For example, PFS, PFS2, time to next subsequent anticancer therapy could be intermediate endpoints that could be informative about the life-expectancy of the patients and are important for treatment decisions. Thus, please consider adding the following bullet: • Surrogate endpoints could be used to complement the added benefit assessment, in particular when such endpoints are labelled in the prescribing information		
Pooyeh Graili	10	287/3	It is misleading because the source of evidence		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Quality HTA			may not be a trial.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	10	281	Cardiovascular diseases are the leading cause of death worldwide: 1. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death		Duplicated comment. The text will be amended
ISPOR – The Professional Society for Health Economics and Outcomes Research	10	284	Core clinical outcomes sets are hardly universal, and many(most) indications may not have these. Suggest providing guidance on what instrument characteristics should be assessed to understand their validity for the specific treatment and indications (e.g., they are fit for the context of use).		Thank you for your comment.
ISPOR – The Professional Society for Health Economics and Outcomes Research	10	284	After the sentence “Specific definitions of outcomes typically used in oncology are provided in Appendix A,” please provide a short list of key outcomes for rare disease ideally coming from horizon scanning of the new technologies expected in the next 5 years.		We consider this out of scope of this guideline.
ISPOR – The Professional Society for Health Economics and Outcomes Research	10	284	In the “Points of attention for the assessment of scoping process” box: Please define “well established” more clearly.		
ISPOR – The	10	287	The source of evidence may not be a trial; we suggest		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Professional Society for Health Economics and Outcomes Research			adding “or the available data” to the end of this sentence.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	10	297	Must a morbid event be irreversible in order to classify as a significant clinical outcome (eg, exacerbations in COPD)?		
ISPOR – The Professional Society for Health Economics and Outcomes Research	10	299	Small point – “rigorously” probably covers the same territory as “rigorously fully” and sounds less redundant.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	10	308	It would be good if the text could encourage use of rigorously validated surrogate outcomes whenever possible for reasons of speed, ethics and costs		
ISPOR – The Professional Society for	10	308	In the “Points of attention for the assessment of scoping process” box, please consider adding: The scoping process should consider evidence on the meaningfulness		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Health Economics and Outcomes Research			of surrogate measures and how they fit within patient treatment expectations based on patients' direct input.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	11	334-336	Please expand on the frameworks that are cited for assessing surrogate outcomes. More specific guidance about the situations in which they are likely to be considered appropriate and acceptable for JCA would be most useful in this document.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	11	323-324	Please clarify whether this means that the HTD should demonstrate the treatment effect in addition to the strength of the association between the surrogate outcome and the patient-centred outcome.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	11	327-328	Regarding the inclusion of immature data: • Please clarify if the requirement to provide data could potentially interfere with the Data Monitoring Committee responsibilities. • This may not be feasible to provide for all outcomes. • How will immature data be used in the JCA process? • Please provide more specific criteria for "immature data," for example, minimum follow-up time. Should the data be blinded? • This could lead to presentation of small samples with limited interpretations.		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
ISPOR – The Professional Society for Health Economics and Outcomes Research	11	313	Please clarify whether both individual patient-level data (IPD) from clinical trials and meta-analysis of trials are necessary for Level 1 evidence – IPD are not likely to co-reside with evidence from trial level meta-analyses.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	11	317	More detail on what constitutes 'consistent' in Level 2 would be helpful. Are multiple observational studies (Level 2) necessarily required in all cases?		
ISPOR – The Professional Society for Health Economics and Outcomes Research	11	331	Section “Uncertainty- Requirements of JCA Reporting” Please reiterate that the assessor should also report evidence on patients’ direct perception of meaningfulness of the surrogate outcome. Please clarify if prior evidence of surrogacy in the same MOA is considered valid or applicable. “An indication of whether or not a patient-centered outcome is likely to be available at a later date.”: This is likely a reference to OS. Is it worth assessing available evidence on the surrogate alongside other available patient-centered outcomes such as HRQoL and Tx satisfaction (most likely available, and these will demonstrate patient perception of tolerability and potential symptom/ HRQoL improvement) while evidence on other patient centered outcome such as OS are awaited?		

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			"In cases for which the association between the surrogate outcome and the final patient-centred outcome has previously been examined but for a different disease stage, population or intervention, the assessment report should consider the implications for the validity of this association in the current population and intervention of interest.": Please clarify if prior evidence of surrogacy in the same MOA is considered valid or applicable.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	12	4.1 Terminology for JCA (Lines 337-345, inclusive of box)	In the spirit of patient centeredness, safety aspect from patients' point of view, patient tolerability, needs to be included. Considering the challenges with causality assessment, efforts need to be made to capture safety events that are less frequently observed including SUSARs using standard terminologies (such as PRO-CTCAE). See more elaboration on this point in the comments below.		Already addressed issue
ISPOR – The Professional Society for Health Economics and Outcomes Research	12	4.1 Terminology for JCA (Lines 337-345, inclusive of box)	The document recommends not to use the term tolerability. However, some symptomatic AE (non-severe but sometimes very bothersome for the patients) are sometimes named under the term "tolerance profile or tolerability". Moreover, FDA and EMA are asking for a more systematic report by patients of symptomatic AE (using the PRO-CTCAE in oncology) in clinical trials, to complement the report by physicians (Safety) but also to describe the tolerability of the new therapy by patients. Repeated studies have shown that for symptomatic AE, there is a disagreement between patients and physicians not in the presence/absence of AE, but in terms of severity and impact on daily life (i.e. physicians tend to		Already addressed issues

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			systematically underreport severity and impact of symptomatic AE, e.g. nausea compared to patients)		
ISPOR – The Professional Society for Health Economics and Outcomes Research	12	4.1 Terminology for JCA (Lines 337-345, inclusive of box)	It may not be necessary to mandate to use one term instead of another. Some of these terms do mean different things and cannot be used interchangeably to many people, such as “adverse event”/ “adverse reaction” vs. “side effect”, and “safety” vs. “tolerability” (search on web and you will see their differences). If the real purpose here is to mandate certain variables to be reported instead of putting restrictions on terminology, providing clear definitions of adverse event and safety might serve better.		We think that multiple terms will be confusing.
ISPOR – The Professional Society for Health Economics and Outcomes Research	12	4.1 Terminology for JCA (Lines 337-345, inclusive of box)	While we agree with avoiding the use of diverse terminologies for the same concept, sometimes terms do differ. Consider that “adverse events are unintended pharmacologic effects that occur when a medication is administered correctly, while a side effect is a secondary unwanted effect that occurs due to drug therapy”.		See previous point
ISPOR – The Professional Society for Health Economics and Outcomes Research	12	342-344 and 371-372	Adverse reactions etc. are commonly part of the Common Technical Documents (CTDs). Avoiding the use of this terminology will be inconsistent with the CTDs. Later in 371-372 it says there are exceptions. More clarification around the use of these safety terms is needed.		CTD is not part of the JCA, and is not binding for the JCA
ISPOR – The Professional Society for Health	12	4.2 Safety: overall and specific adverse events (Lines 346-354,	Along with MS defining their required safety outcomes, both general and specific, it would be helpful if levels of severity of interest for these outcomes (in addition to the normal definition of “serious AE”) are clearly defined as		There is no grading/ranking during the assessment scope (see D4.2 Scoping process)

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Economics and Outcomes Research		inclusive of box)	well.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	12	4.2 Safety: overall and specific adverse events (Lines 346-354, inclusive of box)	Do overall safety results (i.e., all AEs combined) mean % patients with one or more AEs here? If yes, it may still be useful to report incidence of each AE. Assessors may need to see the list to decide which AEs are more important, especially for innovative health technologies.		Yes it does. But listing of all AE will made the JCA report unreadable and must be avoided.
ISPOR – The Professional Society for Health Economics and Outcomes Research	12	354-355	In the “Points of attention for the assessment of scoping process “box: In the spirit of patient-centeredness, one shouldn't discount the patient voice in treatment safety. As such, the term 'patient tolerability', used in this context, should not be excluded.		Already addressed issue
ISPOR – The Professional Society for Health Economics and Outcomes Research	12	368-369	The guideline requires SUSARs (Any suspected unexpected serious adverse reaction) to be reported. A rationale is not given if/how SUSAR reporting could inform relative effectiveness assessments. In principle, SUSARs address pharmacovigilance questions and are a reporting obligation to health authorities by HTD. SUSARs will undergo rigid evaluation by HTD to decide whether a specific event is a new safety signal. A specific event that had been reported as SUSAR in the beginning of a study need not necessarily to be reported at the end of trial following the safety evaluation. In summary the SUSAR evaluation are reflected in the current drug label. Serious adverse events are reported by the HTD in the HTA		Already addressed issue

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			dossiers anyway. Please consider deleting the following sentence: Any suspected unexpected serious adverse reaction (SUSAR) should be reported, even if these are (by definition) not requested during the assessment scoping stage.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	13	373-374	Please clarify if these are ‘drug discontinuations', 'trial discontinuations' or both		It will be updated in the next version.
ISPOR – The Professional Society for Health Economics and Outcomes Research	13	399-400	It should be clear that the patient is not always able to directly provide their perspective (e.g., young children, patients with some cognitive or physical disabilities). In these circumstances, a nonprofessional observer can complete a questionnaire based upon observed manifestations of specific symptoms or impacts (ObsRO) or a professional assessment based upon specific tests that also require an element of judgment to arrive at the score (ClinRO).		We will consider it for the next version of the draft.
Pooyeh Graili Quality HTA	6 & 7	147-177/2	It looks good. However, I do group them as patient-reported outcomes and clinical-reported outcomes (clinician and paraclinical (lab, imaging, etc.) findings). Performance outcomes is misleading because those are partly reported by the patient and partly by the clinicians and healthcare professionals regardless of the type of technology (digital or non-digital). However, we need to keep the third group of outcomes which are achieved by wearables and mobile apps. We		We will clarify the classification for the next version of the draft.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
			may call them 'technology outcomes'. I do not call them performance outcomes because the apps to monitor patients' health are not necessarily measuring performance.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	16	499-509	We acknowledge that distribution-based methods are informative to statistically characterize MIDs. However, for a relevant patient-centered outcome, responder definition should be primarily considered anchor-based methods ideally utilizing patient-reported anchors. Please consider changing the paragraph as follows: MIDs are also frequently estimated using distribution-based methods [51]. In contrast to anchor-based methods, only the overall variability in scores is used in distribution-based methods. Thus, they are criticized as they do not explicitly refer to the meaning of the change for patients (51). Two approaches are most common. The first is based on estimation of Cohen's d. (delete the rest of the sentence in the guidance document)		This will be clarified for the next version of the draft.
ISPOR – The Professional Society for Health Economics and Outcomes Research	16	492-498	In the absence of patient-reported anchors, clinician-reported anchors should be used. Please add the following sentence online as follows: When patient-reported anchors are not available clinician reported anchors could be acceptable.		It will be considered for the next version of the draf.
ISPOR – The Professional Society for Health Economics and Outcomes	16	485-490	The terms MID and MCID are not interchangeable although they are frequently confused and used as if they were the same thing. Suggest clarifying which you are referring to. The definition in Trooster (2011) may be useful: " The minimal effect that would be meaningful to patients is the minimally clinically important difference		This definition is also one of the many definitions of the MCID, which is not the same than the original one. There are never ending debates on terminology about interpretability of scales. We have used the term MID and MCID as they are

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Research			(MCID), the minimal difference that reflects a true improvement (or deterioration) in an outcome is the minimally important difference."		the most prevalent and we will clarify within the document these terms refer to estimate a RD threshold that helps to interpret within patient change as the minimum amount of change that is perceived as such by the patient. We do not think the debate between meaningful and minimal is really helpful to introduce into a practical guideline, because while there is a lot of literature insisting on the "meaningful" aspect, the definition of "what is meaningful" is generally lacking. At least "minimal" is less polysemic: it's the change a patient feels as such, and it is usually the change that is assessed using a PGIC.
ISPOR – The Professional Society for Health Economics and Outcomes Research	16	494	PGIS (impression of severity) may also be used as an anchor		We agree PGISs have sometimes been advocated for defining responder definition according to the patient's perspective, as proposed by the FDA as the end of the 2000s. However, we have not decided to describe this approach for multiple reasons. First, the uptake of this method is largely lower than the use of PGIC, and literature is scarcer about the validity and reliability of this method than the one based on PGIC. Second, while we do agree PGICs are prone to recall bias, there are conceptual issues with the use of PGISs that are frequently overlooked but are fundamental. First, they are a global

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
					measure of the targeted construct (e.g., HRQoL) at one time of measurement. In that regard, they can be considered as a redundant second assessment of the same outcome that the one that is measured by the instrument of interest but with less validity and reliability. It could therefore be argued that, while advocated as such, they are not anchor-based methods as they do not anchor the change in scores to another phenomenon. Rather, they repeat the measure of the same phenomenon. PGIC allows the measure of a different but related construct (i.e., the perceived change by the patient) to the construct of interest. Second, to define a RD using two PGISs, it is the researcher or healthcare professional who must define what is "change" (for example as a change in at least one unit between the two PGISs) before linking it to the change in scores of the instruments of interest. Therefore, the rule used for defining that there is change does not come from an assessment of the perception of the patients on that change. Thus, while they are advocated as such, RD computed by using two PGISs cannot be considered as MIDs according to the patient perspective.
ISPOR – The	8 & 9	217 & 237	Death and mortality are highlighted as outcomes, but many diseases are not fatal or even have "irreversible"		Thank you for your comment.

Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?	HOG response
Professional Society for Health Economics and Outcomes Research			events. Some mention of other important outcomes, eg, pain relief, mental health, and return to work seems warranted.		
ISPOR – The Professional Society for Health Economics and Outcomes Research	21	677-682	While OS, PFS, EFS etc. are certainly relevant to patients, it should be recognized that these are not specifically considered to be patient centered. As per Reeve at al. (2013), patient centered research is "the integration of patients' perspectives about their health with clinical and biological data to evaluate the safety and effectiveness of interventions. Such integration recognizes that health-related quality of life (HRQoL) and how it is affected by disease and treatment complements traditional clinical endpoints such as survival or tumor affected by disease and treatment complements traditional clinical endpoints such as survival or tumor response in cancer."		Thank you for your comment.