ISPOR
The professional society for health economics and outcomes research

Improving healthcare decisions

505 LAWRENCE SQUARE BLVD SOUTH
LAWRENCEVILLE, NJ 08648

P +1-609-586-4981
F +1-609-586-4982

info@ispor.org
www.ispor.org

August 2, 2022

Dear EUnetHTA:

ISPOR – the professional society for health economics and outcomes research - is pleased to respond on behalf of its membership to your Methodological Guidelines consultation "**"D.4.5 Applicability of Evidence" and "D4.6 Validity of Clinical Studies."** We thank you for the opportunity to comment on these draft guidelines.

ISPOR is a scientific and educational society with many of its members engaged in evaluation of health technologies, including pharmaceuticals, medical devices, and other interventions. We have a large membership living and working in 110 countries globally, across a range of disciplines, including health economics, epidemiology, public health, pharmaceutical administration, psychology, statistics, medicine, and more, from a variety of stakeholder perspectives, such as the life sciences industry, academia, research organizations, payers, patient groups, government, and health technology assessment bodies. The research and educational offerings presented at our conferences and in our journals are relevant to many of the issues and questions raised in this request for information.

The response to this consultation was led by members from our Health Science Policy Council, with comments solicited from a number of our membership groups including our Statistical Methods in Health Economics and Outcomes Research Special Interest Group, Meta-Analysis and Network Meta-Analysis Task Force authors, HTA Roundtables, Institutional Council, and Real-World Evidence Steering Committee and RWE Special Interest Group. The attached document provides a synthesis of their comments. We hope they prove useful.

ISPOR would be happy to answer any questions about our response, as well as to participate in any follow-up consultations on the relevant program items mentioned within the report.

Sincerely,

Nancy S. Berg
CEO & Executive Director
ISPOR

EUnetHTA 21 Public Consultation
of D4.5 Applicability of evidence & *D4.6 on Validity of Clinical studies & D5.1 JCA/CA Submission Dossier Template*

Comments should be submitted not *later than 02 August 2022, 23:59 CET*

Please use this form for submitting your comments and share your completed comment form to JCA_Secretariat@zinl.nl prior to the deadline (02 August 2022, 23:59 CET). When submitting your comment form, please include "EUnetHTA 21 – Public Consultation – D4.5, D4.6, D5.2'' in the subject line of your e-mail.
Please carefully read the principles for public consultation here, prior to your review, as these are binding for our process.

We kindly ask you to:

1. Submit one consolidated response per organisation; in a word-file
    a. PDF files will not be accepted;
2. Complete the first table; if this table is not completed, the input will not be considered by EUnetHTA 21;
3. Put each new comment in a new row;
    a. Please be clear about the context of your comment and if possible, provide a suggestion for rewording;
    b. Please consider the HTA Regulation (EU) 2021/2282 when reviewing the document and when you provide comments;
    c. Please consider the corresponding project plan when commenting. Comments that refer to matters out of the scope of the deliverable may not be considered by EUnetHTA 21.
    d. Please do not provide linguistic comments, as the document will undergo language editing prior to finalization;
4. Insert the page number and line/section number on which your comment applies. If your comment relates to the document as a whole, please put **'general'** in this column;
5. Provide a description of your comment as specific as possible and preferably also provide a suggestion for rewording. If you wish to draw our attention to published literature, please supply the full reference;
6. Add rows as needed.

NB: All comments received within the deadline of the consultation and following the correct format will be published on the website, together with the final deliverable. Only comments eligible for consideration will be answered by EUnetHTA 21. The answers will be made publicly available as well. EUnetHTA 21 may decide to rank the comments received on importance.

Please complete this table. If this is not completed, your comments will not be considered.

Name organisation & abbreviation	ISPOR – The Professional Society for Health Economics and Outcomes Research
Country	Headquarters are in the US, but we are a global and international scientific and educational society.
Contact details (name & e-mail address) – *this information will not be published*	Richard Willke – rwillke@ispor.org Kelly Lenahan – klenahan@ispor.org

Sub-deliverable	**Comment from**	**Page number**	**Line/ section number**	**Comment and suggestion for rewording**	**Is your comment an editorial comment?**
	Insert your name and organisation ***Please repeat in each row***	***Insert 'general' if it relates to the whole document*** ***Please don't put 'p' before the number***		***Please insert each new comment in a new row.***	***Please indicate with 'x' if your comment is an editorial comment.***
D4.5 on Applicability of evidence	ISPOR	General		This is a good, very useful document, with clear and specific definitions and very relevant technical information for critical reading and the analysis of scientific literature in the generation of the HTA reports. The contents show an integral and descriptive approach. The document is helpful for the assessment and reporting processes at the national level.	
D4.5 on Applicability of evidence	ISPOR	General		Overall, this document summarizes these important statistical topics in a high level, but some aspects of the discussions may made clearer with more details. In particular, references to good case examples for how multiplicity adjustments are made in protocols (with pre-specified JCA country requirements). or use of sensitivity analyses, would be helpful.	

Please add extra rows as needed.

Sub-deliverable	Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?
D4.5 on Applicability of evidence	ISPOR	General		We do have one significant concern. From reading this document one would get the impression that the only source of evidence in scope is RCTs (despite a brief comment about "various study designs" on l. 133). HTA Regulation (EU) 2021/2282 (35) notes that observational studies can be helpful. And D4.6.1 (validity of clinical studies) includes some discussion of study designs based on RWE, which might range from an external control arm or supplemental survival data to a full retrospective cohort study. However, we found the discussion in D4.6.1 quite limited relative to the attention some other major agencies are paying to RWE. This particular consultation is similarly limited in that respect. Use of non-randomized study designs, primarily based on RWD, creates nuances for each of the 4 main statistical areas covered, introduces other analytical considerations, and suggests special attention to evidence synthesis involving both RCT and RWE results (especially if they are considered different levels of evidence). This would be important additional information, particularly given the proliferation of new treatment innovations with small or precisely defined patient populations, where regulatory approval may have been obtained based on a limited evidence base; this is where JCA could be particularly helpful for member states. Will a future consultation be paying more explicit attention to considerations related to RWE?	
D4.5 on Applicability of evidence	ISPOR	5	102	Member states are required to give "due consideration" to JCA reports, but the paragraph goes on to explain that differences may relate to different member state interpretation of consistency / mismatch between HTD research questions and JCA assessment scope PICO questions. This is helpful context, but it would be helpful to more explicitly spell out the implications of such variation, and in what scenarios this might or might not be appropriate. Distinction and overlap between EUnetHTA guidelines is addressed in the final paragraph of the Introduction.	

Please add extra rows as needed.

Sub-deliverable	Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?
D4.5 on Applicability of evidence	ISPOR	6	130	On line 130 it is stated that this guideline predominantly deals with methodological issues related to inferential statistical analyses. This is a helpful clarification, and could be reflected earlier (or even in the title of the guidance document).	
D4.5 on Applicability of evidence	ISPOR	6	133-135	This document uses “effectiveness” as a common term to describe efficacy, effectiveness and safety (Line 133-135). Given that dossiers may include evidence from both trial and non-interventional, real-world observational study designs, where “effectiveness” is usually considered to mean treatment outcomes in real world, usual care conditions. In addition, sometimes “effectiveness” has been considered a more comprehensive term meaning something more along the lines of “net benefit”. That said, we don’t have a better general term to suggest than “effectiveness” here – terms like “outcome” or “endpoint” don’t work any better in this document - so these considerations may simply bear more explanation here. We can provide references if desired.	
D4.5 on Applicability of evidence	ISPOR	6	133-135	In addition to the definitional point on “effectiveness” above, some of the issues and considerations should be discussed separately with respect to efficacy/effectiveness and safety endpoints. For instance, when a study targets efficacy endpoint, the risk of false positives should be limited, so multiplicity adjustment is necessary when multiple testing to control for type I error rate. However, if a study targets safety endpoint, the economic cost of false negatives could be more significant compared to false positives. Therefore, we want to limit the risk of false negatives to ensure adequate study power to detect safety risks, and we generally are not worried about multiplicity controls for safety endpoint.	
D4.5 on Applicability of evidence	ISPOR	6	169	The statement “the alternative hypothesis is accepted" is technically incorrect. Generally in hypothesis testing, if the p-value is less than the alpha level, we have strong evidence to believe that the null hypothesis is not true,	

Please add extra rows as needed.

Sub-deliverable	Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?
				but it doesn't necessarily mean the evidence is strong enough to believe that the alternative hypothesis is true. Therefore, the statement here should be "the null hypothesis is rejected".	
D4.5 on Applicability of evidence	ISPOR	6-10	3 Multiple Statistical Hypothesis Testing in Individual Clinical Studies (lines 150-258)	The discussion of multiplicity adjustment should be under the context of pre-defined primary, secondary and sensitivity analyses. For instance, one may conduct a sensitivity analysis to assess the treatment effect in a subpopulation. Since the purpose of this sensitivity analysis is to assess the robustness of the primary analysis result and it does not contribute to the main interpretation of the study finding, multiplicity adjustment accounting for this sensitivity analysis is not needed. The examples provided for multiple testing between Line 187-194 may or may not lead to multiplicity issues, depending on the pre-defined study plan on primary, secondary and sensitivity analyses. Therefore, the language of this paragraph here should not be deterministic. And it would benefit from more in-depth discussions to avoid ambiguity.	
D4.5 on Applicability of evidence	ISPOR	6-10	3 Multiple Statistical Hypothesis Testing in Individual Clinical Studies (lines 150-258)	It's worth noting that if the study is for regulatory submission, the multiplicity adjustment can also depend on your regulatory purpose, eg, efficacy (where multiplicity is usually corrected for) vs. safety (where it often is not).	
D4.5 on Applicability of evidence	ISPOR	6-10	3 Multiple Statistical Hypothesis Testing in Individual Clinical Studies (lines 150-258)	In addition to statistical significance and corresponding approaches handling multiplicity problems in different scenarios (e.g., multiple outcomes, multiple time points, multiple treatments, multiple groups, or multiple effect measures), it might be useful to discuss suggestions on how one can utilize clinical significance to guide decision-making in different scenarios.	
D4.5 on Applicability of evidence	ISPOR	7	3 Multiple Statistical Hypothesis	Footnote: For more complex situations, it’s worth noting that the use of simulation studies to explore the type I error control is an overall good strategy to help better	

Please add extra rows as needed.

Sub-deliverable	Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?
			Testing in Individual Clinical Studies (lines 150-258)	understand the multiplicity adjustment issue.	
D4.5 on Applicability of evidence	ISPOR	8	235	One of the requirements for JCA reporting is “Accurate and unambiguous endpoint definitions”. Is this a good opportunity to align with regulatory language and introduce the term “estimands” into HTA reporting?	
D4.5 on Applicability of evidence	ISPOR	8	235	Requirements for appropriate reporting of methods and results in a JCA are concise and straightforward, which can facilitate assessor/co-assessor’s review. However, it is important to separate primary/secondary endpoints rather than combine into multiple endpoints. Moreover, statistical methods for dealing with multiplicity for primary endpoints, and secondary endpoints if applicable should be reported in a JCA	
D4.5 on Applicability of evidence	ISPOR	8	3.2.1 / 3.2.2.	The multiplicity concept seems to be restricted to the context of a single RCT. The guidance should consider other situations where non RCT are needed for the analyses needed for the PICOs for JCA. More broadly, the document should provide some guidance on the PICOs and CER level needed to test across all analyses requested by the JCA. The list of PICOs will determine the testing strategy, which may also vary across the different requests/PICOs from the MS. Should we rather apply a more targeted approach for each single MS (but across PICOs), or have a general approach for each single PICO? This may also require clear directions from JCA on the ordering and importance of each of the PICOs.	
D4.5 on Applicability of evidence	ISPOR	8-9	243-254	It would be helpful to provide further guidance on whether the nominal alpha / alpha spent at interim analysis should be used to analyse all subsequent endpoints / PICOs requested by JCA for the HTA dossier.	
D4.5 on Applicability of evidence	ISPOR	8-9	241, 254, 258	The requirements for reporting are largely related to prespecified statistical planning and transparency - these requirements seem to address statistical reporting rather	

Please add extra rows as needed.

Sub-deliverable	Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?
				than multiplicity specifically. There is also very little practical guidance on methods; references in this section would be helpful.	
D4.5 on Applicability of evidence	ISPOR	11	299	This section (4.2.3) and the next (4.2.4) provide examples with limited discussion of the concepts generally. Additional guidance/references as to how to handle these issues from a conceptual perspective would be helpful.	
D4.5 on Applicability of evidence	ISPOR	13-15	5 Subgroup Analyses in Individual Clinical Studies (lines 341-405)	The considerations of subgroup analyses may be different for superiority test and non-inferiority test. For instance, could the document discuss in the context of non-inferiority test, should the margins be the same within each subgroup, or should different margins be applied depending on subgroup characteristics?	
D4.5 on Applicability of evidence	ISPOR	13-15	5 Subgroup Analyses in Individual Clinical Studies (lines 341-405)	Given the fact that most subgroup analyses do not have sufficient power, in addition to the detailed reports on methods and results from subgroup analyses, it may be useful to present post hoc (or observed) power analyses. Also see and reference the following: Wang et al. Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. *NEJM* 2007; 357:2189-2194.	
D4.5 on Applicability of evidence	ISPOR	13	5.1	A few examples and references in section 5.1 would be helpful, particularly for “variables that represent methodological characteristics of a study” (l. 356).	
D4.5 on Applicability of evidence	ISPOR	14	372-377	It may be worth noting that choosing different cut-offs for a subgroup variable requested by different countries may provide contradicting results, which could in turn merit further analysis.	
D4.5 on Applicability of evidence	ISPOR	14	387-396	Discuss some challenges of subgroup analyses in sample sizes. Could the document discuss the Bayesian approach in subgroup analyses to estimate the subgroup treatment effect? It’s known that Bayesian approach can better handle small and imbalance subgroup sample size to obtain more reliable results (Henderson et al. 2016). Reference: Henderson, N. C., Louis, T. A., Wang, C., & Varadhan, R. (2016). Bayesian analysis of heterogeneous treatment effects for patient-centered outcomes research. *Health Services and Outcomes*	

Please add extra rows as needed.

Sub-deliverable	Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?
				Research Methodology, 16(4), 213-233.	
D4.5 on Applicability of evidence	ISPOR	14,15	378 and 419	The statement "An interaction test is a requirement" should be followed by some discussion on the limitations of the subgroup analyses using such an interaction test. Rightfully, there is mention of the limited power of the test. There should also be mention of the risk of type I error and a wording suggesting that a single p-value for interaction should not be the only tool for identifying or excluding subgroups findings.	
D4.5 on Applicability of evidence	ISPOR	15-16	6 Subgroup Analyses in Evidence Synthesis Studies (lines 406-444)	In some cases, meta-regression is another way of exploring heterogeneity; some discussion of this approach seems merited. This is particularly important in evidence synthesis using direct or indirect treatment comparisons. Meta-regression model focuses more attention on the studies with a lower sampling error and it is able to achieve this by assuming a mixed-effects model. MR model accounts for the deviation from the true overall effect due to sampling error and between-study variance or heterogeneity. Also, you can use one or more variables to predict differences in the true effect sizes.	
D4.5 on Applicability of evidence	ISPOR	16-18	7 Sensitivity Analyses in Individual Studies (lines 445-506)	In addition to sensitivity analyses related to the five attributes of the estimand (i.e., population, treatment, variable (endpoint), intercurrent events and the summary measure), it might be useful to include competing approaches as part of sensitivity analyses. In practice, it is common that multiple models can give undistinguishable goodness-of-fit to the same data set, but different interpretations or conclusions. For example, in evidence synthesis studies, it is common that fixed-effects model can give different conclusions from random-effects model. Reporting results from both approaches. Along with their strengths and limitations, can enhance the summary of evidence generated from systematic reviews and meta-analyses.	
D4.5 on Applicability of evidence	ISPOR	16	459	This section introduces ICEs and missing data as potential type of ICE then proceeds to focus on that. It would be helpful if they provided other examples of ICEs	

Please add extra rows as needed.

Sub-deliverable	Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?
				and the role of sensitivity testing in those cases.	
D4.5 on Applicability of evidence	ISPOR	18	512	Sensitivity analysis could also include tests for something related to the effect of interest like broader inclusion criteria to see if results are consistent or completely opposite like a falsification test.	
D4.5 on Applicability of evidence	ISPOR	19	9 Post Hoc Analyses in Individual Clinical Studies (lines 518-553)	If "supplementary analyses" are provided, it should be specified whether they were pre-specified or post-hoc analyses.	
D4.5 on Applicability of evidence	ISPOR	19	9 Post Hoc Analyses in Individual Clinical Studies (lines 518-553)	While post-hoc analyses are sometimes merited due to unforeseen considerations, it should be noted that there is also potential bias in the selection of which post-hoc analyses are reported and which are not reported.	
D4.6 – Validity of clinical studies	ISPOR		General	Your sections on real-world evidence and non-randomized studies are well-written but in general reflect traditional thinking in these areas. However, recent developments in natural experiments (note the 2021 Nobel Prize in economics) and target trial emulation (note the RCT-DUPLICATE work) have highlighted the potential for valid causal inference with observational data. We also think the value of external validity that real world evidence can contribute to decision-making is understated. While we certainly support the investigation of potential biases via tools like ROBINS-I, we thought your discussion of this area could have been more forward-looking, particularly given a growing need for post-approval evaluations. A fuller explication of this viewpoint can be found in a recent Value in Health Commentary: Berger ML, Crown WC. How Can We Make More Rapid Progress in the Leveraging of Real-World Evidence by Regulatory Decision Makers? *Value in Health*, Volume 25, Issue 2, 167 – 170.	
D4.6 –	ISPOR		General	We would also like to encourage practices, such as study	

Please add extra rows as needed.

Sub-deliverable	Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?
Validity of clinical studies				protocol registration and use of a standard study protocol template, that would improve the reliability and credibility of RWE studies in general. An ISPOR-ISPE special task force report on a standardized RWE study protocol template will be published this fall, while our position on protocol registration can be found here: Orsini LS, Berger M, Crown W, Daniel G, Eichler H-G, Goettsch W, Guerino J, Jonsson P, Lederer NM, Monz B, Mullins D, Schneeweiss S, Van Brunt D, Wang SV, Willke RJ. Improving Transparency to Build Trust in Real-World Secondary Data Studies for Hypothesis Testing—Why, What, and How: Recommendations and a Roadmap from the Real-World Evidence Transparency Initiative. *Value in Health* 2020; 23(9):1128-36 Willke RJ, Wang SV. Registering Study Protocols: Helping RWE Come of Age. *Value & Outcomes Spotlight*. Nov/Dec 2021	
D4.6 – Validity of clinical studies	ISPOR		III. Clinical Study Designs (Lines 235-326)-311/313	Suggestion to change the term 'intervention' to 'exposure' in the context of observational studies	
D4.6 – Validity of clinical studies	ISPOR		IV. Specific strengths, weaknesses, and recommendations regarding different designs (Lines 327-449)-372	How should "effect modifiers" be chosen? Clinicians, comparison of outcome variable by intervention in fitted model, literature search, are there preferred variables by disease area or indication?	
D4.6 – Validity of clinical studies	ISPOR		V. Particularities (Lines 450-658)-459-464	Is a master protocol recommended outside of these three subtypes? In previous work, this was implemented to make definitions and descriptions more consistent and comparable across individual clinical study protocols?	
D4.6 – Validity of clinical	ISPOR		V. Particularities (Lines 450-658)-615	What's the meaning of 'decentralised adjudication'? Should the difference of assessment timelines for certain endpoints (e.g. pfs in oncology study) also be mentioned?	

Please add extra rows as needed.

Sub-deliverable	Comment from	Page number	Line/ section number	Comment and suggestion for rewording	Is your comment an editorial comment?
studies					
D4.6 – Validity of clinical studies	ISPOR		VI. References (Lines 569-790)-630	"observational data from routine healthcare practices, data from registries can be considered as RWD" is there a preference for specific databases or vendors by disease area?	

Please add extra rows as needed.