HTA19

Predicting acceptability of quantitative evidence synthesis for joint clinical assessment submissions: A rating scale

Heldt D[1], Mistry D[2], Ainsworth C[2], Kroep S[3]

[1]OPEN Health HEOR & Market Access, Berlin, DE, [2]OPEN Health HEOR & Market Access, London, UK, [3]OPEN Health HEOR & Market Access, Rotterdam, NL

INTRODUCTION

- The EU regulation on Health Technology Assessment (HTA) becomes active in 2025, providing a permanent framework for joint clinical assessment (JCA) submissions and replacing individual country-specific guidelines that had varying requirements for comparative evidence, including indirect treatment comparisons (ITCs).
- ITCs will be key components in JCA submissions, which require addressing diverse market interests in a single submission, and manufacturers face uncertainty regarding the acceptance of their chosen ITC methodologies.
- This shift has sparked industry discussion, with concerns over how to ensure that comparative evidence will meet the expectations of the assessors across various markets under the new JCA framework.
- A rating system would provide a useful tool to indicate how confident manufacturers can be in the assessors' acceptance of their chosen studies, and analytical approaches, for quantitative evidence synthesis.

OBJECTIVES

The aim of this study was to develop a rating system that could be used to predict the acceptability of data, and methods, used to generate results from ITCs for JCA.

METHODS

- As an initial step, published resources, developed by the HTA Coordination Group, were searched for practical and methodological guidelines on quantitative evidence generation for JCA dossiers.[1,2] These were then reviewed and any recommendations for ITC methodology were extracted. These are presented in this poster.
- As quality of the included studies and their input data is a key component of ITC reliability, an additional step focused on quality was considered to inform the rating scale.
- Criteria to assess individual study quality were selected using Cochrane's risk of bias (RoB) questionnaire[3] alongside additional factors such as sample size, publication type, reporting of treatment effect modifiers (TEMs) and prognostic variable (PVs), study type, outcome assessment method, as well as year and region of study conduct.

RESULTS

- Building on the steps above a rating scale was developed grouping the identified key items into three component subscales:

1. Study-level subscale: Quality of included studies (**Table 1**)
  - Calculated based on factors such as RoB tool outcomes, sample size, and study type.
  - Since some factors may be less important for certain analyses, each was categorised as either universally applicable or situation-specific to reflect its relevance in each use case.

2. Network-level subscale: Exchangeability assumption assessment (**Table 2**)
  - Including study similarity, homogeneity, and consistency.

3. Methodology subscale: Appropriateness of ITC methodology (**Table 3**)
  - Considering the availability of randomised or non-randomised evidence and individual patient data, (IPD) and treatment effect assumptions.

- Within each subscale, items were color-coded (green = recommended or preferred in most situations, orange = use with caution or rigorous justification, pink = not advised), based on expert opinion and interpretation of recommendations from the JCA guidelines.
- Finally, a composite score calculation was developed (**Table 4**).
  - Within each component, a score of four was assigned when all fields were recommended or preferred in most situations (green). Of note, only the assessment items that are applicable for each use case should be chosen for the score calculation.
  - An orange field signaled an outcome or approach requiring caution or rigorous justification (score of 3).
  - A pink field indicated that the approach was not advised (score of 0).
- A total combined score of 12 suggests a high likelihood of acceptance for the analysis method and selected studies. Scores between 9 and 11 indicate that the appropriateness of the methods and study selection may be challenged, while a score below 9 suggests a low likelihood of acceptance.
- The tables in this poster focus on the analytical approaches used for rating scale development, but the practical guidelines for quantitative evidence generation strongly emphasize reporting criteria, which are included in the rating scale (data not shown) and can be shared upon request via email.

Table 1. Rating subscale for study quality assessment

Criteria		Outcome	
In all situations	RoB	Mostly positive	Mostly negative
	Sample size	≥50	<50
	Publication type	Journal article	Conference abstract
	TEM and PV reporting	All identified variables included	Identified variables missing
	Study type	RCT	Non-randomised study
Situation-specific	Outcome assessment	Independent review committee (IRC)	Investigator
	Region of study conduct	Relevant	Not relevant
	Year of study conduct	Relevant	Not relevant

Abbreviations: IRC: Independent review committee; PV: Prognostic variable; RCT: Randomised controlled trial; RoB: Risk of bias; TEM: Treatment effect modifier.

RESULTS

TABLE 2. Rating subscale for assessing study exchangeability

Element		Outcome
Similarity	Assumption holds	Assumption does not hold for all trials
Homogeneity	Assumption holds	Assumption does not hold for all trials
Consistency	Assumption holds (tested with inconsistency models/ node-splitting)	Does not hold for all trials in a closed loop evidence network

TABLE 3. Rating subscale for ITC methodology assessment

Situation		Approach	
Effect	Large number of studies	Random effects (RE)	Fixed effect (FE)
	Small number of studies	RE with weakly informative prior FE	
	Only two studies	FE or RE	
Non-randomised data	Anchored	Individual patient data (IPD) available for non-randomised study, test for shifted null hypothesis included	IPD not available for non-randomised study
		Simulated treatment comparison (STC)	STC with non-identity link function
		Multilevel network meta-regression (ML-NMR)	Matching-adjusted indirect comparison (MAIC)
	Unanchored	IPD available for all studies (for all co-variates: confounders, PVs, TEMs and prespecified in statistical analysis plan [SAP]), test for shifted null hypothesis included	Without IPD
		STC	
		MAIC	
Randomised data	Unanchored	IPD available	No IPD available
	Anchored approaches (proportional hazards [PH] assumption holds [Time-to-event [TTE] outcomes])	MAIC or STC (IPD not available for all studies), if TEMs and PVs are clinically justified, strategy for selection was prespecified in SAP, and the effect size is large enough that it couldn't be induced by missing TEMs alone	
		ML-NMR	
		Bucher	Bucher including multi-arm trial
			Bucher is inappropriate to estimate treatment effects or consistency tests, when RE models have been used.
		Network meta-analysis (NMA)	RE for one or more comparison
		NMR	STC (IPD available for all studies
		MAIC (IPD available for all studies)	
	Anchored approaches (PH assumption violated [TTE outcomes])	NMA of restricted mean survival time, fractional polynomial NMA, cubic spline models, parametric survival curves, piecewise exponential models	Other approaches were applied

Abbreviations: FE: Fixed effects; IPD: Individual patient data; MAIC: Matching-adjusted indirect comparison; ML-NMR: Multilevel network meta-regression; NMA: Network meta-analysis; PH: Proportional hazards; PV: Prognostic variable; RE: Random effects; SAP: Statistical analysis plan; STC: Simulated treatment comparison; TEM: Treatment effect modifier; TTE: Time-to-event.

TABLE 4. Composite rating scale score calculation

Study quality (score)	Exchangeability (score)	Methodology (score)	Total score
All applicable fields green: recommended or preferred in most situations (4)	All applicable fields green: recommended or preferred in most situations (4)	All applicable fields green: recommended or preferred in most situations (4)	12
Any orange: use with caution or rigorous justification (3)	Any orange: use with caution or rigorous justification (3)	Any orange: use with caution or rigorous justification (3)	≥9 and <12
		Any pink: not advised (0)	<9

CONCLUSIONS

- A rating system for quantitative evidence synthesis will provide a useful tool to predict the acceptability of ITC results in JCA dossiers and to anticipate and mitigate potential barriers at an early stage.
- Among the methods, NMA or Bucher ITCs are favored when feasible. Population-adjustment methods are highly uncertain and require rigorous justification if used, but should be avoided when IPD is available.
- A new emphasis has emerged on the comprehensive identification of potential treatment TEMs and PVs, alongside the reporting criteria for methods. Although a specific methodology for TEM/PV identification is not yet recommended, clear reporting and a thorough consideration of their impact, and any necessary ITC adjustments, are essential.
- Other factors influencing acceptability, such as unmet need or the availability of evidence within an indication, remain unexplored, and future research is needed to incorporate these in a holistic rating system. The rating criteria may be applied less strictly to rare indications or those with high unmet need, where anchored comparisons or reliance solely on RCT evidence may not be feasible.
- Finally, the proposed rating system must be tested on published dossiers to validate and refine its approach, ensuring its practicality and robustness in future assessments.

REFERENCES

1. Member State CG on HTA. (2024). Methodological Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons; **2.** Member State CG on HTA. (2024). Practical Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons; **3.** Centre for Reviews and Dissemination. Systematic Reviews. CRD's guidance for undertaking reviews in health care, 2009.

ACKNOWLEDGEMENTS

We would like to thank Claire Simons and Elisabeth Fenwick for taking the time to review this poster.

Presented at: ISPOR Europe 2024

Sponsored by OPEN Health's Scientific Office

SOCA1_Poster_v0.2_JS_301024.indd 1 31/10/2024 12:05