Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Epistemological Paradox of EU-HTA Joint Clinical Assessments: Certainty Without Judgment, Evidence Without Context

Mondher Toumi

mto@inovintell.com

Aix-Marseille Université

Szymon Jarosławski

Aix-Marseille Université

Lylia Chachoua

Clever-Access

Laurent Boyer

Aix-Marseille Université

Wael Fourati

Clever-Access

Pascal Auquier

Aix-Marseille Université

Olga Burzynska

Clever-Access

Borislav Borissov

Medical University of Sofia

Claude Dussart

University Claude Bernard Lyon 1

Research Article

Keywords: Health technology assessment, Epistemic certainty, EU-HTA Regulation, Relative effectiveness, Scientific judgment, Pragmatic trials

Posted Date: August 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10667201/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The European Union Health Technology Assessment Regulation (EU-HTA Regulation 2021/2282) mandates Joint Clinical Assessments (JCAs) that must describe “the degree of certainty of the relative effectiveness” while simultaneously prohibiting “any value judgment.” This article examines the fundamental epistemological tensions embedded in this regulatory framework. We analyse how EU-HTA attempts to assess certainty, an inherently epistemic concept requiring interpretive synthesis, while claiming to operate in a judgment-free, decontextualised space. Drawing on philosophical distinctions between epistemic and ontic uncertainty, we demonstrate that integrating multiple heterogeneous dimensions of uncertainty into a singular degree of certainty necessarily involves scientific judgment and implicit contextual assumptions. We further examine the conflation of efficacy and effectiveness in EU-HTA’s approach, showing how reliance on explanatory randomised controlled trials (RCTs) fundamentally limits the assessment of real-world relative effectiveness despite regulatory terminology suggesting otherwise. We conclude that these tensions are not incidental but structural: the aspiration to judgment-free, context-free certainty assessment is internally inconsistent. Transparency, we argue, would be better served by explicitly structured and documented judgment, as exemplified by frameworks such as GRADE, than by the current narrative reporting format.

1. Introduction

The EU-HTA Regulation (2021/2282), fully applicable from January 12, 2025, represents a landmark harmonization effort in European health technology assessment [1]. At its core lies a fundamental epistemological challenge: Joint Clinical Assessments (JCAs) must provide "a scientific analysis of the relative effects of the health technology" and assess "the degree of certainty of the relative effects, taking into account the strengths and limitations of the available evidence" [1] (Article 9). Simultaneously, the regulation explicitly prohibits JCAs from containing "any value judgment or judgments on the overall clinical added value" [1] (Article 9(1)).[2, 3]

While Article 9 itself prohibits only “any value judgment or judgments on the overall clinical added value,” the true scope of this prohibition is set out far more expansively in Recital 28 of the Regulation, which provides that:

“The joint clinical assessment report should be factual and should not contain any value judgment, ranking of health outcomes, conclusions on the overall benefit or clinical added value of the assessed health technology, any position on the target population in which the health technology should be used, or any position on the place the health technology should have in the therapeutic, diagnostic or preventive strategy” [1] (Recital 28).

Read together, Article 9 and Recital 28 establish a prohibition of remarkable breadth. The assessment must remain strictly “factual”; it may not rank health outcomes, may not reach conclusions on overall benefit or clinical added value, may not take any position on the target population in which the technology should be used, and may not take any position on the place of the technology within the

therapeutic, diagnostic, or preventive strategy. Considered in isolation, Article 9 conveys an incomplete and potentially misleading picture of the regulatory intent. It is Recital 28 that reveals the full ambition of the framework: not merely to exclude a narrow class of evaluative preferences, but to insulate the joint clinical assessment from the contextual and comparative determinations that ordinarily give clinical evidence its meaning.

Therein lies the tension this article examines. The HTAR, and the subsequent HTA Coordination Group guidance, seek to separate assessment from context by prohibiting positions on the target population and on the technology's place in the therapeutic strategy, the ranking of outcomes, and conclusions on clinical added value, while simultaneously requiring a conclusion on the degree of certainty of the relative effects. The resulting framework aspires to a judgment-free, decontextualized assessment of evidence. Yet certainty assessment may itself require precisely the contextual judgments that the framework seeks to exclude.

This mandate creates a philosophical paradox. How can one reliably assess certainty, a concept fundamentally rooted in epistemic evaluation, without exercising scientific judgment? How can heterogeneous evidence across multiple outcomes and dimensions of uncertainty be integrated into a coherent degree of certainty while maintaining a judgment-free stance? And how can such assessments claim to evaluate "relative effectiveness" when the evidentiary base consists predominantly of explanatory trials designed to establish efficacy under ideal conditions rather than effectiveness under real-world circumstances?

These questions are not merely academic. They strike at the operational core of a regulatory system that will shape pharmaceutical access for 450 million European citizens. The tensions between certainty and uncertainty, between judgment-free assessment and evaluative synthesis, and between efficacy and effectiveness represent more than methodological challenges; they embody fundamental contradictions in how contemporary health technology assessment conceptualizes the relationship between evidence and decision-making.

This article examines these tensions through a philosophical lens, analyzing the conceptual foundations of certainty assessment, the epistemological requirements for integrating heterogeneous uncertainty, and the pragmatic implications of narrative reporting formats for public decision-making. We argue that EU-HTA's current framework contains irresolvable contradictions that undermine its stated objectives of transparency, reproducibility, and decision-support.

2. Philosophical Foundations: Epistemic Certainty and Scientific Uncertainty

2.1 The Ontic-Epistemic Distinction

To understand the certainty assessment challenge within the EU-HTA framework, it is necessary to recognize that health technology assessment (HTA) navigates three primary domains of uncertainty: ontic, epistemic, and interpretive.

First, ontic (or ontological) uncertainty refers to the inherent randomness or indeterminacy in the world itself—aleatory uncertainty that exists independently of our state of knowledge. In clinical medicine, this manifests as the irreducible variability in individual treatment responses, disease progression, and biological outcomes. Distinguishing this from epistemic uncertainty—which stems from limitations in our knowledge and could theoretically be reduced with better information, enhanced methods, or additional data—is widely recognized as a cornerstone of rigorous decision modeling [4–7].

While this distinction is philosophically elegant, it becomes operationally ambiguous in HTA. For instance, imprecision in effect estimates—such as wide confidence intervals—may reflect epistemic uncertainty (e.g., small sample sizes or methodological limitations), ontic uncertainty (e.g., genuine heterogeneity in treatment effects across populations), or, as is frequently the case, a combination of both [4, 6, 8]. The EU-HTA's mandate to assess "the degree of certainty" implicitly targets epistemic concerns—asking how confident we should be in our knowledge claims—yet the regulation provides no framework for disentangling these epistemic and ontic components.

Finally, the framework must account for interpretive uncertainty. This form of uncertainty arises from the fundamental need for scientific judgment when translating evidence into system-specific decisions. Even when evidence is statistically robust, HTA bodies must interpret what that evidence means within a particular healthcare context, including making judgments about clinical relevance, generalizability, appropriate comparators, and acceptable trade-offs [8–11].

Crucially, interpretive uncertainty persists even when epistemic uncertainty has been substantially reduced; it is therefore structural, not residual. Consequently, it remains exceedingly difficult to assess certainty without addressing this interpretive layer—particularly concerning constructs like the Minimum Clinically Important Difference (MCID). Research suggests that MCID is often an imperfect proxy for clinical relevance, as estimates show wide discordance across populations and methods, requiring clinical and patient-centered interpretation rather than serving as a fixed universal threshold [12–15]. Ultimately, explicit judgment is required at all levels to meaningfully assess uncertainty.

2.2 Certainty Versus Uncertainty: A Non-Symmetrical Relationship

The scientific method is fundamentally oriented toward quantifying and managing uncertainty. Researchers express findings probabilistically, use confidence intervals to bound parameter estimates, conduct sensitivity analyses to explore assumption-dependence, and employ Bayesian methods to formally update beliefs given evidence [16]. The framework is one of continuous probability distributions, where uncertainty is the primitive concept and certainty represents an unattainable ideal.

EU-HTA, however, requires assessment of certainty, a concept that inverts this relationship. As Wittgenstein argued in “On Certainty”, epistemic certainty is not merely the absence of doubt but a distinct psychological and philosophical state [17]. Certainty implies confidence that "the true effect is within a particular range" [18], a fundamentally different construct than probabilistic uncertainty about parameter values.

This asymmetry creates a conceptual mismatch. Scientists address uncertainty through continuous quantification (probability distributions, confidence intervals, posterior distributions). Regulatory bodies seek certainty through categorical or ordinal classifications. The translation from one framework to the other is neither straightforward nor neutral, it requires interpretive judgment about threshold effects, clinically meaningful differences, and the sufficiency of evidence for decision-making.

2.3 GRADE: The Dominant Framework for Certainty Assessment

The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach represents the most widely adopted framework for assessing certainty in evidence [19]. GRADE evaluates evidence across multiple domains, risk of bias, inconsistency, indirectness, imprecision, and publication bias, to assign certainty ratings (high, moderate, low, very low) to effect estimates for each outcome [19].

Critically, GRADE explicitly acknowledges that "judgment is involved during the evidence assessment" and that "overall certainty reflects if and how much concerns about the domains matter" [20]. GRADE does not eliminate judgment; it structures and makes it transparent. As the GRADE handbook notes, "the great merit of GRADE is not that it ensures reproducible judgments but that it requires explicit judgment that is made transparent to users" [21].

EU-HTA's prohibition of judgment thus creates an immediate problem: how to assess certainty without the methodological framework explicitly designed for that purpose? The regulation's guidance documents do not specify an alternative certainty assessment methodology. Without GRADE or an equivalent structured approach, how can assessors reliably and reproducibly evaluate "the degree of certainty" across diverse evidence types?

3. The Challenge of Judgment-Free Assessment

3.1 The Impossibility of Value-Neutral Evaluation

EU-HTA's prohibition of value judgments in JCAs reflects a fundamental misunderstanding of scientific assessment. As recent critiques have noted, "assessors must determine comparative effectiveness, which requires evaluating study design, inclusion criteria, outcomes, and their impact on validity and statistical precision. Prohibiting judgment leads to flawed assessments" [22].

The breadth of this prohibition is confirmed, and indeed widened, by Recital 28. By requiring that the report remain strictly "factual" and by barring any "ranking of health outcomes," the Regulation forecloses precisely the comparative weighting that the integration of heterogeneous evidence into a single degree of certainty would seem to demand. By prohibiting any "position on the target population" and any position on the technology's "place…in the therapeutic, diagnostic or preventive strategy," it excludes the very clinical and decision-context considerations that determine whether a given level of uncertainty is tolerable. The prohibition is therefore not a narrow safeguard against arbitrary preference; it is a structural commitment to decontextualized evaluation. That commitment stands in direct tension with the contextual character of certainty assessment, because the judgments the Regulation seeks to remove are not external additions to certainty assessment but constitutive elements of it.

HTA work explicitly concludes that HTA "by and large is a process of value judgments" present in framing questions, choosing endpoints and comparators, synthesizing evidence, and economic analysis; these judgments are "basic elements of the HTA process," not external add-ons [23, 24]. For advanced therapies, value judgments about endpoints, comparators, and distributional concerns are seen as unavoidable, and making them explicit is proposed as a way to improve HTA accountability [24].Consider the apparently straightforward task of assessing risk of bias. Tools like the Cochrane Risk of Bias 2 (RoB 2) require evaluators to judge whether: - Randomization processes were adequate - Deviations from intended interventions occurred - Outcome data were missing in ways that could bias results - Outcome measurement was appropriate and consistent - Selective reporting occurred.

Even though RoB 2 is formally structured, it still requires subjective classification:

- RoB 2 overall judgment is reached only after answering signalling questions and then assigning categories of low/some concerns/high risk, an evaluative step [25, 26].
- Empirical studies show low inter-rater reliability for RoB 2 overall judgments and several domains, even among experienced reviewers, and describe the tool as "difficult and demanding," requiring calibration and training [27].
- Detailed case experience reports that reviewers had to develop their own cut-points and rules (e.g. for missing data thresholds, interpreting patient-reported outcomes) and worried that judgments rested on assumptions rather than firm evidence [28].
- Meta-research finds that many non-Cochrane reviews using RoB 2 do not justify judgments, underscoring how much discretion is involved [29].

GRADE's treatment of indirectness makes the role of judgment explicit. Users must first define a target PICO (population, intervention, comparator, outcome) that reflects the decision context, then judge how far each study's PICO deviates and whether these differences are likely to change effect size [19, 30].

Indirectness includes population and intervention differences, but also outcomes such as surrogate endpoints, which are "not themselves important" but presumed to reflect patient-important outcomes; certainty is rated down depending on how convincing this linkage is judged to be [19]. For test accuracy,

GRADE guidance similarly stresses that accuracy data are typically surrogates for patient-important outcomes; unless authors link test performance to downstream outcomes, certainty for decisions should be rated down for indirectness [31].

Further GRADE work on values and preferences shows that even judging the importance of outcomes involves assessing risk of bias and indirectness in preference studies, again using PICO-based judgments [32]. Introductory GRADE summaries emphasise that certainty ratings always integrate study design, RoB, inconsistency, indirectness, imprecision, and publication bias in a contextual, deliberative process, not a value-free algorithm [33].

3.2 The Integration Challenge: From Multiple Uncertainties to Singular Certainty

Perhaps the most profound challenge in certainty assessment is integration across heterogeneous dimensions. Consider a typical JCA scenario:

- Outcome A (mortality): Low risk of bias, precise estimate, directly relevant, inconsistent across trials - Outcome B (quality of life): High risk of bias, imprecise estimate, indirectly measured, consistent across trials - Outcome C (adverse events): Moderate risk of bias, very imprecise, directly relevant, suspected publication bias

How should these be synthesized into "the degree of certainty of the relative effects"? Each outcome has different certainty profiles. The overall assessment requires:

1. Weighting the relative importance of outcomes (inherently value-laden)
2. Trading off different dimensions of uncertainty (risk of bias vs. imprecision vs. indirectness)
3. Determining thresholds for when uncertainty becomes "serious" enough to downgrade certainty
4. Synthesizing across outcomes to reach an overall judgment

None of these steps can be performed algorithmically. Each requires scientific and contextual judgment. As GRADE explicitly recognizes, "when the body of evidence is intermediate with respect to a particular factor, the decision about whether a study falls above or below the threshold for downgrading depends on judgment" [21].

EU-HTA's mandate to perform these integrative assessments while prohibiting judgment and forbidding outcome weighting, creates a logical impossibility. The very concept of "degree of certainty" presupposes a synthesized evaluation but weighting and synthesis requires judgment. Without the ability to weigh outcomes and apply scientific judgment, assessing certainty becomes impossible.

3.3 The Implicit Context Problem

EU-HTA claims to provide "decontextualized" assessments that Member States can then adapt to local contexts. This conceptual separation is illusory [34, 35]. Certainty is inherently context-dependent in multiple ways:

Clinical context dependency

Whether a surrogate endpoint provides "indirect" evidence depends on understanding of disease pathophysiology in specific populations [36–38]. Whether effect size uncertainty is "serious" depends on the magnitude of effect needed for clinical meaningfulness [39]. Whether an adverse event impacts uncertainty depend of the nature of the adverse event, the seriousness, severity, the duration and the reversibility. Mortality and transitory nausea may not have the same impact on relative effectiveness. These require clinical context [40].

Decision context dependency

The "sufficiency" of evidence depends on decision stakes [39, 40]. For life-threatening conditions with no alternatives, lower certainty evidence may suffice. For elective interventions with established alternatives, higher certainty may be required. These are context-specific judgments [41, 42].

Methodological context dependency

The severity of publication bias depends on the incentive structures in specific therapeutic areas [30, 31]. The concern about unblinding depends on the objectivity of outcome measurement in particular disease contexts, as well as the models used [43, 44].

By claiming to provide context-free certainty assessments, EU-HTA obscures the inevitable contextual assumptions embedded in its evaluations. The result is not truly decontextualized assessment, rather, it is assessment with implicit, unstated contextual assumptions that Member States cannot readily identify or modify [45].

4. The Efficacy-Effectiveness Conflation

4.1 Explanatory Versus Pragmatic Trials

The distinction between efficacy and effectiveness is fundamental to evidence interpretation. Efficacy trials (explanatory trials) determine whether an intervention produces expected results under ideal circumstances, highly controlled conditions designed to maximize internal validity [46]. Effectiveness trials (pragmatic trials) measure the degree of beneficial effect under real-world clinical settings, prioritizing external validity and generalizability [46].

As Porzsolt and colleagues demonstrate, "it is impossible to design a trial which can detect efficacy and effectiveness simultaneously. The RCTs describe what we may expect in health care, while the pragmatic controlled trials describe what we really observe" [47]. The PRECIS-2 framework formalizes this continuum across nine domains including participant eligibility, recruitment, setting, intervention flexibility, and follow-up intensity [48].

Critically, most regulatory trials are highly explanatory: - Strict inclusion/exclusion criteria select homogeneous populations - Standardized intervention protocols prevent optimization - Placebo controls create artificial comparisons - Protocol-mandated follow-up ensures compliance - Specialized centers provide expert care delivery.

Such designs maximize internal validity but inherently limit external validity. The effect observed in these idealized conditions (efficacy) systematically differs from the effect achievable in routine practice (effectiveness).

4.2 EU-HTA's Terminology: Relative Effectiveness or Comparative Efficacy?

The EU-HTA Regulation uses the term "relative clinical effectiveness" [1], yet the evidentiary base consists almost entirely of explanatory RCTs submitted for regulatory authorization [45, 49–51]. This creates a fundamental disconnect.

Pre-authorization RCTs are designed to demonstrate efficacy for regulatory approval, not effectiveness for health technology assessment [49, 52, 53]. They answer: "Does this intervention work under controlled conditions with selected patients?" not "How well does this intervention perform in routine clinical practice with typical patients?" [52, 53].

The terminology "relative effectiveness" misleadingly implies real-world performance assessment [49, 50]. More accurately, EU-HTA performs comparative efficacy assessment, comparing the performance of technologies under idealized research conditions, which frequently suffer from poor external validity [54–56]. The extrapolation from efficacy to effectiveness requires assumptions about:

- How trial populations relate to target populations
- How protocol-driven care relates to routine care
- How specialized settings relate to typical settings
- How monitored adherence relates to real-world adherence [54–57].

None of these extrapolations are trivial, and all require judgment about external validity [50, 55–59]—precisely the contextual judgment EU-HTA structurally prohibits.

4.3 The Measurement Problem: Ratios or Comparisons?

The term "relative effectiveness" carries additional ambiguity. In epidemiology, "relative" typically implies a ratio measure, relative risk, relative rate, hazard ratio [60–62]. These express proportional relationships between intervention and comparator.

However, EU-HTA JCAs do not primarily assess ratio metrics. They perform comparative assessments, structured comparisons across multiple outcomes, often including absolute effect measures, responder

analyses, and clinical significance evaluations [51, 63–65]. The output is not a "relative effectiveness ratio" but a multidimensional comparative profile [66, 67].

This semantic confusion matters. The terminology "degree of relative effectiveness" suggests a summarizable, quantitative construct (e.g., "intervention X is 30% more effective than comparator Y"). The actual output is a complex, multidimensional narrative describing comparative performance across diverse outcomes with varying certainty [6, 45]. These are fundamentally different information structures with different decision-support properties.

5. Certainty Assessment Without GRADE: The Methodological Gap

5.1 The Absence of Standardized Approaches

GRADE is the internationally recognized standard for certainty assessment, endorsed by over 100 organizations globally [21]. It provides: - Explicit criteria for evaluating evidence domains - Structured processes for rating up or down - Standardized output formats (Summary of Findings tables) - Transparent documentation of judgments

EU-HTA guidance documents do not mandate GRADE use. Neither do they specify an alternative structured approach. The Implementing Regulation (EU 2024/1381) establishes procedural rules but not methodological standards for certainty assessment [68].

This creates a significant gap. Without a standardized framework: Reproducibility is compromised, different assessors may reach different certainty judgments for identical evidence, Transparency is limited, the basis for certainty judgments cannot be systematically documented, Comparability across JCAs is impossible, no common metric or framework exists Quality control is hindered, no clear standards exist against which to evaluate assessment quality

The result is likely to be heterogeneous, difficult-to-reproduce certainty assessments that vary based on assessor background, national traditions, and therapeutic area norms rather than systematic methodology.

5.2 The ROBINS-I Exception and Its Implications

One notable exception exists: when using ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) to assess non-randomized studies, evidence can start at high certainty if bias assessment is favorable [20]. This represents implicit GRADE adoption, ROBINS-I was designed specifically for GRADE integration.

The fundamental insight—that study design determines initial certainty, and specific quality features subsequently modify it—is GRADE's core logic. Attempting to assess certainty without this logic risks either: 1. Crude design-based stratification ignoring quality variation 2. Ad hoc, non-standardized quality judgments 3. Implicit, unacknowledged GRADE-like reasoning

The first undermines assessment validity, the second sacrifices reproducibility, and the third creates conceptual incoherence by using GRADE logic while claiming not to.

6. Narrative Format and Decision-Making Utility

6.1 The Structure of JCA Reports

JCA reports follow a standardized template including: health problem characterization, technology description, research questions and scope, methods, results, and discussion [68]. Evidence synthesis and certainty assessment are presented primarily through narrative text, supplemented by evidence tables.

This format contrasts with GRADE's Summary of Findings (SoF) tables, which present: - Each critical/important outcome in rows - Certainty rating (high/moderate/low/very low) per outcome - Absolute and relative effect estimates with confidence intervals - Footnoted explanations for certainty judgments

The SoF format enables at-a-glance comprehension: decision-makers immediately see which outcomes have high-certainty evidence of important effects versus which have low-certainty evidence or uncertain effects. The tabular structure facilitates comparison and supports systematic decision analysis.

6.2 Narrative Limitations for Public Decision-Making

Narrative certainty descriptions face several challenges for decision support:

Information density: Narrative text obscures key information in prose paragraphs. Extracting certainty assessments for specific outcomes requires reading potentially dozens of pages. This increases cognitive load and reduces accessibility.

Non-standardized language: Without controlled vocabularies, different JCAs may describe similar certainty levels differently ("some uncertainty," "moderate confidence," "limited evidence"). This hinders cross-JCA comparison.

Incomplete documentation: Narrative format may omit explicit certainty judgments for some outcomes, particularly secondary outcomes. The absence of documentation creates ambiguity, was certainty assessed and found moderate, or simply not assessed?

Poor machine-readability: Narrative text is difficult to parse computationally. Systematic analysis across JCAs (e.g., patterns in certainty ratings by therapeutic area) requires manual extraction rather than automated processing.

Weak decision support: Decision frameworks like multi-criteria decision analysis (MCDA) or cost-effectiveness modeling require explicit certainty ratings as inputs. Narrative descriptions cannot easily be converted to the quantitative or ordinal certainty measures these frameworks need.

6.3 The Paradox of Transparency

EU-HTA emphasizes transparency as a core principle. Yet narrative certainty assessment may reduce rather than enhance transparency by:

1. Obscuring judgments: Tabular GRADE ratings force explicit judgment documentation. Narrative descriptions can obscure whether judgments were made and on what basis.
2. Complicating review: Stakeholders reviewing draft JCAs face difficulty identifying specific certainty judgments to comment on when these are embedded in continuous narrative.
3. Impeding accountability: Post-hoc evaluation of whether certainty was appropriately assessed requires extracting judgments from narrative rather than examining explicit ratings.

The result is a transparency paradox: the format intended to provide comprehensive information may actually make certainty judgments less accessible, less comparable, and less accountable than structured alternatives.

7. Limitations of This Analysis

This analysis has focused primarily on conceptual and epistemological tensions. We have not empirically examined actual JCA reports (as of this writing, the first JCAs report is not available). Future research should evaluate whether operational practice resolves, exacerbates, or transcends the epistemological challenges identified here.

Additionally, we have focused on certainty assessment for clinical effectiveness. EU-HTA JCAs also assess safety, where similar issues arise but with different risk-benefit trade-offs. The certainty requirements for safety evaluation deserve separate detailed analysis.

Finally, our critique should not be interpreted as opposition to European HTA cooperation. The goals of avoiding duplicative assessment, improving efficiency, and enhancing evidence-based decision-making are laudable. Our analysis aims to strengthen EU-HTA by identifying foundational challenges that, if unaddressed, may undermine its long-term success.

8. Conclusion

The EU-HTA Regulation represents an ambitious attempt to harmonize clinical evidence assessment across diverse European health systems. At its core lies a profound epistemological challenge: how to assess certainty, an inherently evaluative, judgment-dependent, context-sensitive construct, while claiming to provide judgment-free, decontextualized scientific analysis. Recital 28 makes the ambition explicit, requiring a strictly “factual” report stripped of value judgment, outcome ranking, conclusions on clinical added value, and any position on the target population or therapeutic place. Yet the more comprehensively the framework excludes context, the more sharply it exposes the contradiction: certainty cannot be characterized without the very judgments the Regulation forbids.

Our analysis demonstrates that this challenge may be irresolvable in its current formulation. The integration of heterogeneous uncertainty dimensions into coherent certainty assessments necessarily requires scientific judgment. The assessment of whether evidence suffices for conclusions inevitably depends on clinical and decision context. The extrapolation from explanatory trial efficacy to real-world effectiveness inherently involves assumptions beyond the data.

EU-HTA's reluctance to acknowledge these realities, its insistence on judgment-free assessment, context-free evaluation, and terminology that conflates efficacy with effectiveness, creates conceptual incoherence at the foundation of the new regulatory system. Without standardized certainty assessment frameworks, without explicit acknowledgment of necessary judgments, without clear distinction between what trials measure and what policies need, EU-HTA risks producing assessments that are neither scientifically rigorous nor decision-useful.

The path forward requires honest engagement with the epistemological realities of evidence synthesis. Certainty assessment cannot be value-free or judgment-free. It can, however, be structured, transparent, and consistent in how judgments are made and documented. Frameworks like GRADE show how to achieve this. EU-HTA's long-term credibility and utility depend on embracing similar methodological clarity rather than maintaining the fiction of judgment-free certainty assessment.

The broader lesson extends beyond EU-HTA. Health technology assessment worldwide struggles with similar tensions between scientific objectivity and evaluative necessity, between neutrality and utility, between evidence and decision. Resolving these tensions requires not more sophisticated methods for hiding judgment, but more sophisticated epistemological foundation framework. Until health technology assessment embraces this reality, it will continue to face the paradoxes examined in this analysis.

Declarations

Author Contribution

Conceptualization [M.T.].; writing—original draft preparation, [M.T.]; writing—review and editing, [S.J., L.C., L.B., W.F., P.A., O.B., B.B., C.D.]. All authors have read and agreed to the published version of the manuscript. All authors have read and agreed to the published version of the manuscript.

References

1. Regulation, E. U. (2021). 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU. *Official Journal of the European Union L*, *458*, 1–32.
2. Gsteiger, S., Bucher, H., Ryan, J., & Ruof, J. (2024). Technology Assessment vs. Technology Appraisal—How to Strengthen the Science/Value Dichotomy with EU HTA? *Journal of Market Access & Health Policy*, *12*, 369–377. https://doi.org/10.3390/jmahp12040028

3. Schuster, V. (2024). EU HTA Regulation and Joint Clinical Assessment—Threat or Opportunity? *Journal of Market Access & Health Policy*, *12*, 100–104. https://doi.org/10.3390/jmahp12020008
4. Grimm, S., Pouwels, X., Ramaekers, B., Wijnen, B., Knies, S., Grutters, J., & Joore, M. (2019). Development and Validation of the TRansparent Uncertainty ASsessmenT (TRUST) Tool for Assessing Uncertainties in Health Economic Decision Models. *Pharmacoeconomics*, *38*, 205–216. https://doi.org/10.1007/s40273-019-00855-9
5. Cruz, I. R., Troffaes, M., & Sahlin, U. (2022). A suggestion for the quantification of precise and bounded probability to quantify epistemic uncertainty in scientific assessments. *Risk Analysis*, *42*, 239–253. https://doi.org/10.1111/risa.13871
6. Vreman, R., Strigkos, G., Leufkens, H., Schünemann, H., Mantel-Teeuwisse, A., & Goettsch, W. (2022). Addressing uncertainty in relative effectiveness assessments by HTA organizations. *International Journal of Technology Assessment in Health Care*, 38. https://doi.org/10.1017/s026646232100177x
7. Yazdi, M., Kabir, S., & Walker, M. (2019). Uncertainty handling in fault tree based risk assessment: State of the art and future perspectives. *Process Safety and Environmental Protection*, *131*, 89–104. https://doi.org/10.1016/j.psep.2019.09.003
8. Trowman, R., Powers, A., & Ollendorf, D. (2021). Considering and communicating uncertainty in health technology assessment. *International Journal of Technology Assessment in Health Care*, 37. https://doi.org/10.1017/S0266462321000453
9. Hogervorst, M., Vreman, R., Heikkinen, I., et al. (2023). Uncertainty management in regulatory and health technology assessment decision-making on drugs: guidance of the HTAi-DIA Working Group. *International Journal of Technology Assessment in Health Care*, 39. https://doi.org/10.1017/S0266462323000375
10. Vreman, R., Naci, H., Goettsch, W., Mantel-Teeuwisse, A., Schneeweiss, S., Leufkens, H., & Kesselheim, A. (2020). Decision Making Under Uncertainty: Comparing Regulatory and Health Technology Assessment Reviews of Medicines in the United States and Europe. *Clinical Pharmacology and Therapeutics*, *108*, 350–357. https://doi.org/10.1002/cpt.1835
11. Regier, D., Pollard, S., McPhail, M., et al. (2022). A perspective on life-cycle health technology assessment and real-world evidence for precision oncology in Canada. *NPJ Precision Oncology*. https://doi.org/10.1038/s41698-022-00316-1. 6.
12. Draak, T., De Greef, B. D., Faber, C., & Merkies, I. (2019). The minimum clinically important difference: which direction to take. *European Journal of Neurology*, *26*, 850–855. https://doi.org/10.1111/ene.13941
13. Rossi, M. J., Brand, J., & Lubowitz, J. H. (2023). Minimally Clinically Important Difference (MCID) Is a Low Bar. *Arthroscopy*, *39*(2), 139–141. https://doi.org/10.1016/j.arthro.2022.11.001
14. Migliorini, F., Maffulli, N., Memminger, M., Simeone, F., Rath, B., & Huber, T. (2024). Clinical relevance of patient-reported outcome measures in patients who have undergone total hip arthroplasty: a systematic review. *Archives of Orthopaedic and Trauma Surgery*, *144*, 4907–4916. https://doi.org/10.1007/s00402-024-05579-w

15. Deckey, D. G., Verhey, J. T., Christopher, Z., Gerhart, C., Clarke, H., Spangehl, M., & Bingham, J. (2022). Discordance Abounds in Minimum Clinically Important Differences in THA: A Systematic Review. *Clinical Orthopaedics and Related Research*, *481*, 702–714. https://doi.org/10.1097/CORR.0000000000002434
16. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). CRC.
17. Wittgenstein, L. (1969). In Anscombe, von G. H. Wright, & D. Paul (Eds.), *On certainty (G. E. M*. Blackwell. & G. E. M. Anscombe, Trans.).
18. Hultcrantz, M., Rind, D., Akl, E. A., Treweek, S., Mustafa, R. A., Iorio, A., Alper, B. S., Meerpohl, J. J., Murad, M. H., Ansari, M. T., Katikireddi, S. V., Östlund, P., Tranæus, S., Christensen, R., Gartlehner, G., Brozek, J., Izcovich, A., Schünemann, H., & Guyatt, G. (2017). The GRADE Working Group clarifies the construct of certainty of evidence. *Journal of Clinical Epidemiology*, *87*, 4–13. https://doi.org/10.1016/j.jclinepi.2017.05.006
19. Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., DeBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., & Schünemann, H. J. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, *64*(4), 383–394. https://doi.org/10.1016/j.jclinepi.2010.04.026
20. Schünemann, H. J., Cuello, C., Akl, E. A., Mustafa, R. A., Meerpohl, J. J., Thayer, K., Morgan, R. L., Gartlehner, G., Kunz, R., Katikireddi, S. V., Sterne, J., Higgins, J. P. T., Guyatt, G., & GRADE Working Group. (2019). GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. *Journal of Clinical Epidemiology*, *111*, 105–114. https://doi.org/10.1016/j.jclinepi.2018.01.012
21. GRADE Working Group (2024). GRADE handbook. Retrieved from https://gradepro.org/handbook/.
22. Toumi, M., Falissard, B., Jouini, A., Aballéa, S., Boyer, L., & Auquier, P. (2025). EU-HTA Guidance for Clinical Validity: Misconceptions and Flawed Processes. *Journal of Market Access & Health Policy*, *13*(3), 33. https://doi.org/10.3390/jmahp13030033
23. Hofmann, B., Cleemput, I., Bond, K., Krones, T., Droste, S., Sacchini, D., & Oortwijn, W. (2014). Revealing and acknowledging value judgments in health technology assessment. *International Journal of Technology Assessment in Health Care*, *30*(6), 579–586. https://doi.org/10.1017/S0266462314000671
24. Gonçalves, E. (2020). Advanced therapy medicinal products: value judgement and ethical evaluation in health technology assessment. *The European Journal of Health Economics*, *21*(3), 311–320. https://doi.org/10.1007/s10198-019-01147-x
25. Nejadghaderi, S. A., Balibegloo, M., & Rezaei, N. (2024). The Cochrane risk of bias assessment tool 2 (RoB 2) versus the original RoB: A perspective on the pros and cons. *Health Science Reports*, *7*(6), e2165. https://doi.org/10.1002/hsr2.2165
26. Flemyng, E., Moore, T. H., Boutron, I., Higgins, J. P. T., Hróbjartsson, A., Nejstgaard, C. H., & Dwan, K. (2023). Using Risk of Bias 2 to assess results from randomised controlled trials: guidance from

Cochrane. *BMJ Evidence-Based Medicine*, *28*(4), 260–266. https://doi.org/10.1136/bmjebm-2022-112102

27. Minozzi, S., Cinquini, M., Gianola, S., Gonzalez-Lorenzo, M., & Banzi, R. (2020). The revised Cochrane risk of bias tool for randomized trials (RoB 2) showed low interrater reliability and challenges in its application. *Journal of Clinical Epidemiology*, *126*, 37–44. https://doi.org/10.1016/j.jclinepi.2020.06.015
28. Crocker, T. F., Lam, N., Jordão, M., Brundle, C., Prescott, M., Forster, A., Ensor, J., Gladman, J., & Clegg, A. (2023). Risk-of-bias assessment using Cochrane's revised tool for randomized trials (RoB 2) was useful but challenging and resource-intensive: observations from a systematic review. *Journal of Clinical Epidemiology*, *161*, 39–45. https://doi.org/10.1016/j.jclinepi.2023.06.015
29. Babić, A., Barcot, O., Višković, T., Šarić, F., Kirkovski, A., Barun, I., & Puljak, L. (2024). Frequency of use and adequacy of Cochrane risk of bias tool 2 in non-Cochrane systematic reviews published in 2020: Meta-research study. *Research Synthesis Methods*, *15*(3), 430–440. https://doi.org/10.1002/jrsm.1695
30. Guyatt, G. H., Wang, Y., Eachempati, P., Iorio, A., Murad, M. H., Hultcrantz, M., Chu, D. K., Florez, I. D., Hemkens, L., Agoritsas, T., Yao, L., Vandvik, P., Montori, V., & Brignardello-Petersen, R. (2025). Core GRADE 4: rating certainty of evidence—risk of bias, publication bias, and reasons for rating up certainty. *Bmj*, 389. https://doi.org/10.1136/bmj-2024-083864
31. Schünemann, H., Mustafa, R., Brożek, J., Steingart, K., Leeflang, M., Murad, M., Bossuyt, P., Glasziou, P., Jaeschke, R., Lange, S., Meerpohl, J., Langendam, M., Hultcrantz, M., Vist, G., Akl, E., Helfand, M., Santesso, N., Hooft, L., Scholten, R., & Guyatt, G. (2020). GRADE guidelines: 21 part 2. Inconsistency, Imprecision, publication bias and other domains for rating the certainty of evidence for test accuracy and presenting it in evidence profiles and summary of findings tables. *Journal of clinical epidemiology*. https://doi.org/10.1016/j.jclinepi.2019.12.021
32. Zhang, Y., Alonso-Coello, P., Guyatt, G. H., Yepes-Nuñez, J. J., Akl, E. A., Hazlewood, G., & Schünemann, H. J. (2018). Assessing the certainty of evidence in the importance of outcomes or values and preferences—Risk of bias and indirectness. *Journal of Clinical Epidemiology*, *111*, 94–104. https://doi.org/10.1016/j.jclinepi.2018.01.013. GRADE Guidelines: 19.
33. Prasad, M. (2023). Introduction to the GRADE tool for rating certainty in evidence and recommendations. *Clinical Epidemiology and Global Health*, *25*, 101484. https://doi.org/10.1016/j.cegh.2023.101484
34. Németh, B., Goettsch, W., Kristensen, F., Pinyazhko, O., Huić, M., Tesar, T., Atanasijevic, D., Lipska, I., & Kaló, Z. (2020). The transferability of health technology assessment: the European perspective with focus on central and Eastern European countries. *Expert Review of Pharmacoeconomics & Outcomes Research*, *20*, 321–330. https://doi.org/10.1080/14737167.2020.1779061
35. Christensen, R., Ciani, O., Manyara, A. M., & Taylor, R. S. (2023). Surrogate Endpoints: A Key Concept in Clinical Epidemiology. *Journal of clinical epidemiology*. https://doi.org/10.1016/j.jclinepi.2023.111242. 111242.

36. Kemp, R., & Prasad, V. (2017). Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? *BMC Medicine*, 15. https://doi.org/10.1186/s12916-017-0902-9
37. Grigore, B., Ciani, O., Dams, F., Federici, C., De Groot, S., Möllenkamp, M., Rabbe, S., Shatrov, K., Zemplényi, A., & Taylor, R. S. (2020). Surrogate Endpoints in Health Technology Assessment: An International Review of Methodological Guidelines. *Pharmacoeconomics*, *38*, 1055–1070. https://doi.org/10.1007/s40273-020-00935-1
38. Mittal, A., Kim, M. S., Dunn, S., Wright, K., & Gyawali, B. (2024). Frequently asked questions on surrogate endpoints in oncology-opportunities, pitfalls, and the way forward. *eClinicalMedicine*, 76. https://doi.org/10.1016/j.eclinm.2024.102824
39. Zeng, L., Brignardello-Petersen, R., Hultcrantz, M., Siemieniuk, R., Santesso, N., Traversy, G., Izcovich, A., Sadeghirad, B., Alexander, P., Devji, T., Rochwerg, B., Murad, M., Morgan, R., Christensen, R., Schünemann, H., & Guyatt, G. (2021). GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings. *Journal of clinical epidemiology*. https://doi.org/10.1016/j.jclinepi.2021.03.026
40. Schünemann, H., Neumann, I., Hultcrantz, M., Brignardello-Petersen, R., Zeng, L., Murad, M. H., Izcovich, A., Morgano, G. P., Baldeh, T., Santesso, N., Cuello, C. G., Mbuagbaw, L., Guyatt, G., Wiercioch, W., Piggott, T., De Beer, H., Vinceti, M., Mathioudakis, A., Mayer, M., & Akl, E. (2022). GRADE Guidance article 35: Update on rating imprecision for assessing contextualized certainty of evidence and making decisions. *Journal of clinical epidemiology*. https://doi.org/10.1016/j.jclinepi.2022.07.015
41. Castanon, A., Sloan, R., Arocha, L., & Ramagopalan, S. (2024). EU HTA Joint Clinical Assessment: are patients with rare disease going to lose out? *Journal of Comparative Effectiveness Research*, 13. https://doi.org/10.57264/cer-2024-0052
42. Chávarri, A. T., Sacks, A. A., & Val, S. (2025). Contextual aspects of Health Technology Assessment in the new EU framework. *The European Journal of Public Health*. https://doi.org/10.1093/eurpub/ckaf161.227. 35.
43. Brożek, J., Canelo-Aybar, C., Akl, E., Bowen, J., Bucher, J., Chiu, W., Cronin, M., Djulbegovic, B., Falavigna, M., Guyatt, G., Gordon, A., Boon, M. H., Hutubessy, R., Joore, M., Katikireddi, V., LaKind, J., Langendam, M., Manja, V., Magnuson, K., & Schünemann, H. (2020). GRADE Guidelines 30: the GRADE approach to assessing the certainty of modeled evidence—An overview in the context of health decision-making. *Journal of clinical epidemiology*, *129*, 138–150. https://doi.org/10.1016/j.jclinepi.2020.09.018
44. Ciani, O., Grigore, B., & Taylor, R. S. (2022). Development of a framework and decision tool for the evaluation of health technologies based on surrogate endpoint evidence. *Health Economics*, *31*, 44–72. https://doi.org/10.1002/hec.4524
45. Johns, A., Andrade, A., Johal, S., & Ryan, J. (2025). EU Joint Clinical Assessment: A Framework for Optimising Use with Cost-Effectiveness Decision-Making. *Journal of Market Access & Health Policy*,

13. https://doi.org/10.3390/jmahp13040052

46. Gartlehner, G., Hansen, R. A., Nissman, D., Lohr, K. N., & Carey, T. S. (2006). *Criteria for distinguishing effectiveness from efficacy trials in systematic reviews (Technical Reviews* (Vol. 12No.). Agency for Healthcare Research and Quality.
47. Porzsolt, F., Rocha, N. G., Toledo-Arruda, A. C., Thomaz, T. G., Moraes, C., Bessa-Guerra, T. R., Leão, M., Migowski, A., da Silva, A. R. A., & Weiss, C. (2015). Efficacy and effectiveness trials have different goals, use different tools, and generate different messages. *Pragmatic and Observational Research*, *6*, 47–54. https://doi.org/10.2147/POR.S89946
48. Loudon, K., Treweek, S., Sullivan, F., Donnan, P., Thorpe, K. E., & Zwarenstein, M. (2015). The PRECIS-2 tool: Designing trials that are fit for purpose. *Bmj*, *350*, h2147. https://doi.org/10.1136/bmj.h2147
49. Ali, M. S., Prieto-Alhambra, D., Lopes, L. C., Ramos, D., Bispo, N., Ichihara, M. Y., Pescarini, J., Williamson, E., Fiaccone, R., Barreto, M. L., & Smeeth, L. (2019). Propensity Score Methods in Health Technology Assessment: Principles, Extended Applications, and Recent Advances. *Frontiers in Pharmacology*, 10. https://doi.org/10.3389/fphar.2019.00973
50. Makady, A., Van Veelen, A., Jónsson, P., Moseley, O., d'Andon, A., De Boer, A., Hillege, H., Klungel, O., & Goettsch, W. (2017). Using Real-World Data in Health Technology Assessment (HTA) Practice: A Comparative Study of Five HTA Agencies. *Pharmacoeconomics*, *36*, 359–368. https://doi.org/10.1007/s40273-017-0596-z
51. Sarri, G., Vinals, L., Leisle, L., Chau, I., Smalbrugge, D., Lucassen, K., & Jemiai, Y. (2025). Mapping methods gaps between EU joint clinical assessments and local health technology assessment decision-making: an environmental scan of guidance in select EU markets and harmonization challenges. *Journal of Comparative Effectiveness Research*, 14. https://doi.org/10.57264/cer-2024-0240
52. Verleye, L., Bmj, R. M., & Eldridge, S. (2019). Pragmatic trials: lab meets bedside. *British Journal of Dermatology*, *181*. https://doi.org/10.1111/bjd.17889
53. Le-Rademacher, J., Gunn, H. J., Yao, X., & Schaid, D. (2023). Clinical Trials Overview: From Explanatory to Pragmatic Clinical Trials. Mayo Clinic proceedings, 98 8, 1241–1253. https://doi.org/10.1016/j.mayocp.2023.04.013
54. Kennedy-Martin, T., Curtis, S., Faries, D., Robinson, S., & Johnston, J. (2015). A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials*, 16. https://doi.org/10.1186/s13063-015-1023-4
55. Rothwell, P. (2005). External validity of randomised controlled trials: to whom do the results of this trial apply? *Lancet*, *365 9453*, 82–93. https://doi.org/10.1016/s0140-6736(04)17670-8
56. Rothwell, P. (2006). Factors That Can Affect the External Validity of Randomised Controlled Trials. *PLoS Clinical Trials*, *1*. https://doi.org/10.1371/journal.pctr.0010009
57. Averitt, A., Ryan, P., Weng, C., & Perotte, A. (2021). A conceptual framework for external validity. *Journal of biomedical informatics*. https://doi.org/10.1016/j.jbi.2021.103870. 103870.

58. Moore, F. T., Grunch, B. H., Miller, L., & Musacchio, M. J. (2021). *Towards a More Simplified Approach for Evaluating Strength of Evidence in Health Technology Assessments* (p. 13). Cureus. https://doi.org/10.7759/cureus.16528
59. Trapani, D., Tay-Teo, K., Tesch, M., Roitberg, F., Sengar, M., Altuna, S., Hassett, M., Genazzani, A., Kesselheim, A., & Curigliano, G. (2022). Implications of Oncology Trial Design and Uncertainties in Efficacy-Safety Data on Health Technology Assessments. *Current Oncology*, *29*, 5774–5791. https://doi.org/10.3390/curroncol29080455
60. Argyropoulos, C. (2021). Comments on the clinical meaningfulness of a treatment's effect on a time-to-event variable [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2108.13575
61. Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine*, 21. https://doi.org/10.1002/sim.1201
62. Lim, J., Farhat, I., Douros, A., & Panagiotoglou, D. (2022). Relative effectiveness of medications for opioid-related disorders: A systematic review and network meta-analysis of randomized controlled trials. *Plos One*, 17. https://doi.org/10.1371/journal.pone.0266142
63. Julian, E., Gianfrate, F., Solà-Morales, O., Mol, P., Bergmann, J., Salmonson, T., Hebborn, A., Grande, M., & Ruof, J. (2022). How can a joint European health technology assessment provide an 'additional benefit' over the current standard of national assessments? *Health Economics Review*, 12. https://doi.org/10.1186/s13561-022-00379-7
64. Chassagnol, F., Marcelli, G., Wagle, J., Giuliani, G., Traub, D., Schaub, V., & Ruof, J. (2020). Review of Relative effectiveness assessments (REAs) of pharmaceuticals at the European network for health technology assessment(EUnetHTA): A first step towards a consolidated European perspective on comparative effectiveness & safety? Health policy. https://doi.org/10.1016/j.healthpol.2020.06.013
65. Kisser, A., Knieriemen, J., Fasan, A., Eberle, K., Hogger, S., Werner, S., Taube, T., & Rasch, A. (2021). Towards compatibility of EUnetHTA JCA methodology and German HTA: a systematic comparison and recommendations from an industry perspective. *The European Journal of Health Economics*, *23*, 863–878. https://doi.org/10.1007/s10198-021-01400-2
66. Van Beekhuizen, S., Che, M., Monfort, L., Hashim, M., Azough, A., Kubitz, N., Griffin, A., & Price, M. (2025). Indirect Treatment Comparisons in EUnetHTA Relative Effectiveness Assessments: Learnings and Recommendations for the Implementation of EU Joint Clinical Assessments. *PharmacoEconomics Open*, *9*, 597–609. https://doi.org/10.1007/s41669-025-00575-1
67. Chubak, J., Rutter, C., Kamineni, A., Johnson, E. A., Stout, N., Weiss, N., Doria-Rose, V. P., Doubeni, C. A., & Buist, D. (2013). Measurement in Comparative Effectiveness Research. *American journal of preventive medicine*, *44*, 513–519. https://doi.org/10.1016/j.amepre.2013.01.006
68. European Commission (2024). Commission Implementing Regulation (EU) 2024/1381 of 23 May 2024 laying down, pursuant to Regulation (EU) 2021/2282 on health technology assessment, procedural rules for the interaction during, exchange of information on, and participation in, the preparation and update of joint clinical assessments of medicinal products for human use at Union level. Official Journal of the European Union, L 1381.