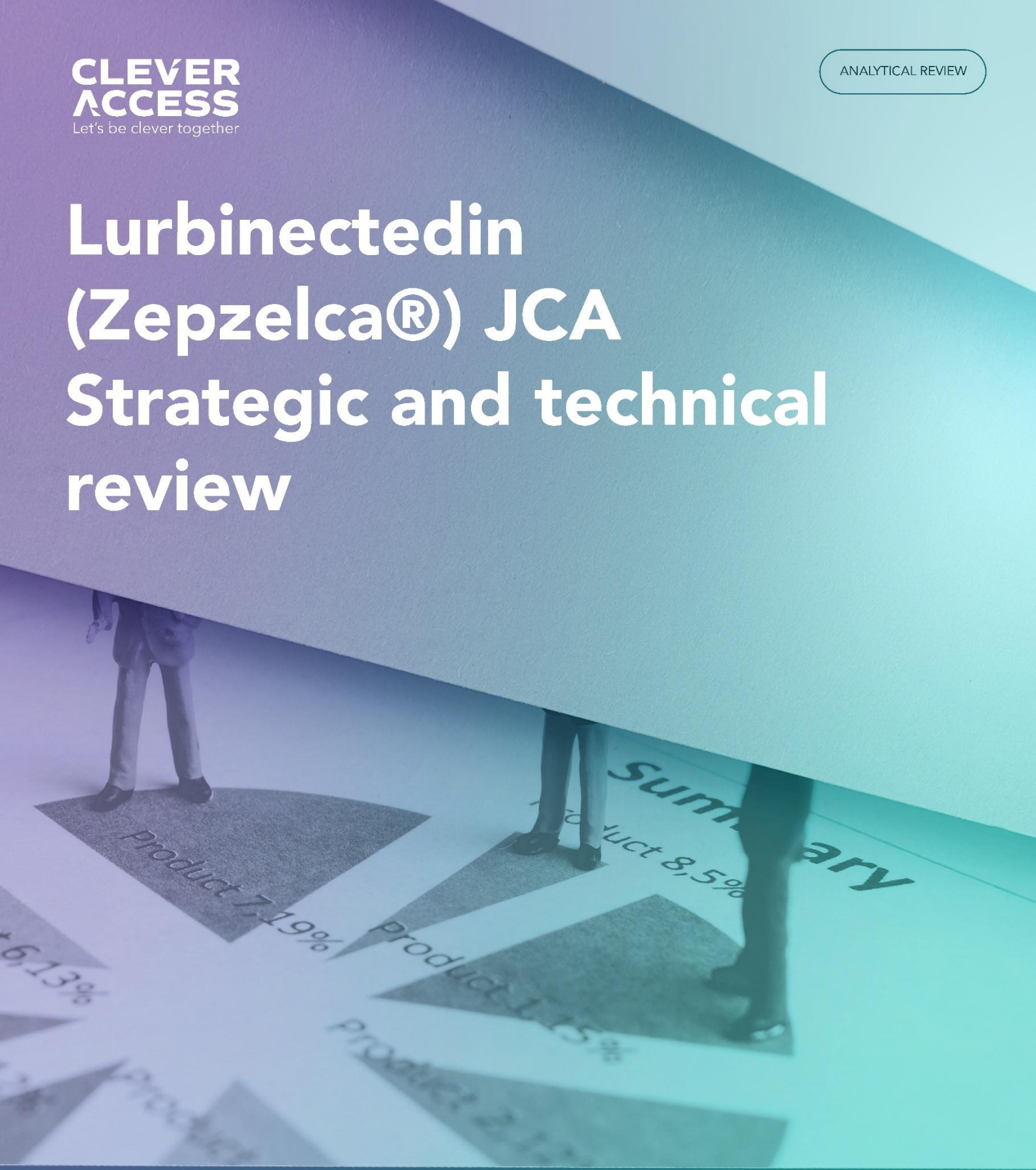
CLEVER ACCESS
Let's be clever together

ANALYTICAL REVIEW

Lurbinectedin (Zepzelca®) JCA Strategic and technical review



ABOUT REPORT

Deep methodological, evidentiary and legal analysis.

AUTHORS


Mondher Toumi


Lylia Chachoua


Aleksandra Caban


Anna Kapuśniak

Published by MAS

CONTENTS

Contents

EXECUTIVE SUMMARY

Lurbinectedin (Zepzelca®)

Joint Clinical Assessment

Executive summary of the deep methodological, evidentiary and legal analysis

Case	JCA-MP-2024-16 — Lurbinectedin (Zepzelca®), Pharma Mar S.A.; first-line maintenance of extensive-stage small cell lung cancer
Assessor / Co-assessor	IQWiG, Germany (B. Wieseler) / INFARMED, Portugal (R. Moura)
Legal basis	Regulation (EU) 2021/2282 (HTAR); Implementing Regulation (EU) 2024/1381
Procedural status	Report v1.0; endorsed by the HTACG on 22 June 2026; EC procedural review 1 July 2026
Evidence base	One randomised, open-label phase III trial (IMforte / GO43104); one population, one PICO
Register	Executive summary (updated version as of 3 August 2026) — the granular factual analysis is provided in the companion technical review below

1. Headline conclusion

This is the cleanest structural case among the early Joint Clinical Assessments: a single population, a single PICO, and a single randomised controlled trial providing direct head-to-head evidence. Lurbinectedin plus atezolizumab is compared with atezolizumab monotherapy as first-line maintenance for extensive-stage small cell lung cancer, based on the IMforte trial. There is no single-arm data, no matching-adjusted comparison, and no network, the structure the JCA framework was principally designed for.

The verdict is that the difficulty here is not structure but interpretation. The assessment turns on a single decisive choice: which data cut of the trial to treat as authoritative. At the earlier, multiplicity-controlled cut favoured by the developer, overall survival is statistically significant (hazard ratio 0.73). At the more mature cut the assessors designate as their main data source, overall survival is no longer nominally significant (hazard ratio 0.81, nominal p=0.0605). The same trial therefore supports either a positive or an equivocal reading, depending on the cut, and the report, consistent with the framework, never resolves that tension into a graded statement of certainty. The most important act in the assessment is interpretive, and it is not labelled as such.

This raises a major issue from a statistical epistemological perspective. The trial is designed within the Neyman-Pearson framework. Under this framework, the trial is built to deliver a single response, the one reported by the HTD. Any additional analysis, with or without a computed p value, even if very low, such as 0.000001, is only hypothesis-generating. The only valid results that allow inference are those used and reported by the assessor. If the assessor wanted to use additional information to inform the appraisal as decisive, they should use the Fisher framework. In the Fisher framework, a p value of 0.06 is considered as good as 0.495 for concluding superiority.

A subtler concern is the report's handling of the mature-cut p-value. The assessors describe it as "nominal" but do not make explicit what that entails: at a data cut outside the sequential testing strategy, no alpha is allocated, so the p-value carries no confirmatory inferential status and is, at most, hypothesis-generating. The report therefore rests part of its reading on a figure whose inferential limits it does not state. It also mixes inferential paradigms, applying a Neyman-Pearson logic of pre-specified error control to set aside the developer's confirmatory analysis while interpreting the mature-cut result in an informal Fisherian manner. That combination is

characteristic of the assessor's national methodological tradition but is not shared across HTA decision frameworks. The effect, whether intended or not, is to steer the reader towards the less favourable reading without a transparent statement of the inferential basis for doing so.

This points to a governance requirement rather than a charge against individuals. The JCA should state and then adhere to the inferential framework it uses to evaluate evidence; where more than one framework could apply, it should select and justify that choice a priori, before the trial data are available to the assessors. Choosing the governing framework, or the authoritative data cut, after the analyses have been seen makes the assessment post hoc: on its own terms, it can then be only hypothesis-generating, not confirmatory, and should not be presented as a basis for inference. Pre-specification is the ordinary safeguard against exactly this hazard, and it is one the framework already requires of developers.

2. Compliance with the binding methodology

The assessment team's conduct is compliant and rigorous. It followed the Annex template, set a single-PICO scope under Article 8(6), documented carer and clinical-expert involvement under Articles 8(6) and 11(4), and applied an outcome-level risk-of-bias assessment (using RoB 1 within a treatment-policy framework) separately for each data cut. It scrutinised the trial's late amendment to its statistical stopping rule, designated the mature data cut as most relevant for health technology assessment, and declined to include analyses it deemed uninterpretable. This is exactly the independent, methodologically assertive review the guidance expects. However, the report does not record that the European regulator examined the same stopping-rule amendment and concluded that all the changes were operational rather than data-driven, nor that the interim result would also have passed the O'Brien–Fleming boundary the amendment replaced. Where the report is open to challenge is not its procedural conduct but its consistency with evidence-based medicine, as section 5 develops: the same evidentiary standard is not applied to the two data cuts, and documented uncertainty is never converted into a graded degree of certainty.

3. How the assessors handled the developer's evidence

The assessors issued a single request for additional information with two items — the first subsequent anti-cancer therapy by arm, and the participant disposition and reasons for discontinuation during the induction phase, where a discrepancy in patient counts raised a selection-bias concern. The developer provided the requested data and an updated disposition flowchart, and the assessors record that these points were fully addressed. Separately, at the factual-accuracy stage, the developer confirmed that certain quality-of-life subgroup analyses in the dossier were incorrect and submitted corrected versions, which were handled outside the body of the report. The more consequential disagreements are methodological rather than factual: which data cut is authoritative, whether the late change to the alpha-spending function is legitimate, and whether the immune-related adverse-event and patient-reported toxicity analyses are interpretable.

4. Certainty versus uncertainty

As in the other early cases, the report documents the factors affecting certainty for each outcome, including population generalisability (patients with poor performance status and brain metastases were excluded), the open-label design, the shift to nominal p-values at the mature cut, and unequal observation windows for safety, but it does not convert these into an explicit, graded certainty rating. There is no high/moderate/low/very-low classification, and GRADE is not used. The gap is especially consequential here, though the word "borderline" describes the p-value rather than the effect: between the two data cuts, the hazard ratio moves only from 0.73 (95% CI 0.57 to 0.95) to 0.81 (95% CI 0.65 to 1.01), and the upper confidence bound shifts only from 0.95 to 1.01. Thus,

the apparent change from a significant to a non-significant survival benefit is largely an artefact of the p-value crossing 0.05 (0.0174 to 0.0605) rather than a change in the estimated benefit. Two facts make that adjudication lean towards a maintained benefit: the mature-cut analysis is, by design, hypothesis-generating rather than confirmatory because its alpha was spent at the interim; and the control arm was partly contaminated by subsequent lurbinectedin, with 22 control-arm patients (9.1%) receiving the experimental agent itself, which biases the intention-to-treat estimate towards the null. That contamination is not a feature of the mature cut alone: all 22 switches had occurred by the first cut, and the count was identical at both dates. What separates the two analyses is not the number of patients exposed to non-protocol therapy but the length of time they accrued under it. Given what the difference means, a probable but only modestly attenuated benefit rather than none, the report withholds precisely the graded certainty conclusion. Two further factors, developed in the technical review, point the same way: the loss of nominal significance is consistent with dilution of the intention-to-treat contrast by control-arm access to the experimental agent (the developer's analysis addressing this was set aside as "not considered relevant"), and the assessors accepted a weighted-average hazard ratio under non-proportional hazards for other endpoints while treating a threshold-crossing survival p-value as decisive.

5. Evidence-based medicine (EBM) compliance

The report invokes “international standards of evidence-based medicine” and, because the evidence is a randomised trial, satisfies EBM’s core preference for randomised comparison. The selective application is evident in what is excluded rather than retained and graded: the immune-related adverse-event analysis is excluded for methodological flaws and incompleteness, and the patient-reported toxicity (PRO-CTCAE) results are excluded as not interpretable. EBM would more typically retain such analyses at low certainty and bound their interpretation. The exclusions are defensible on their own terms, but they remove from view precisely the tolerability evidence most relevant to maintenance therapy in patients who, by definition, are responding to prior treatment.

6. Implicit judgement and reader influence

The report reaches no conclusion on added value, and so respects the assessment/appraisal boundary. Yet the decision to prefer the mature, less favourable data cut, the criticism of the alpha-spending amendment, and the exclusion of the toxicity analyses are consequential interpretive choices that steer the national reader towards a cautious reading of a therapy whose earlier, multiplicity-controlled analysis was positive. Each is defensible as scientific judgement; none crosses into appraisal. But together they demonstrate that even in the framework’s cleanest structural case, the decisive influence on the national reader is exercised through implicit methodological choices rather than an explicit, graded verdict.

7. Learnings for future JCA report development

- **Pre-specify which data cut is authoritative for HTA.** The single most consequential decision in this assessment — the choice between an earlier significant cut and a mature non-significant one — was made after the fact and without a graded certainty statement. It was also made without recording that the regulator had already examined the sole objection to the earlier cut and had found the protocol changes to be operational rather than data-driven.
- **Explicitly grade certainty.** A borderline survival signal is exactly the situation in which a reproducible uncertainty-to-certainty step is indispensable.
- **Retain and bound tolerability analyses rather than excluding them.** For maintenance therapy, immune-related and patient-reported toxicities are decision-critical and should be presented with caveats, not removed.

- **Trace stakeholder influence.** A carer independently observed the survival attenuation at the mature cut and the unfavourable quality-of-life signals; the report does not record how these observations shaped it.
- **Use Joint Scientific Consultation.** No consultation preceded this case; the data-cut and stopping-rule questions could have been settled in advance.

The consistent bottom line: even with the cleanest possible evidence structure, the lurbinectedin JCA still leaves its most important judgement — how much confidence a borderline survival benefit deserves — unstated. That confirms, from the opposite end of the evidence spectrum to the single-arm cases, that the missing uncertainty-to-certainty step is a structural feature of the framework.

STRATEGIC REVIEW

Lurbinectedin (Zepzelca®)

Joint Clinical Assessment

Strategic lessons from the framework's cleanest structural case

Subject	Joint Clinical Assessment of lurbinectedin (Zepzelca®) plus atezolizumab as first-line maintenance for extensive-stage small cell lung cancer
Framework	Regulation (EU) 2021/2282 (HTAR); methodological guidance of the HTACG
Significance	A single-RCT, single-PICO case — the structure the JCA was designed for; the tensions here are purely interpretive
Register	Strategic/governance — the granular factual analysis is held in the companion technical review

OVERVIEW

Executive Summary

IN BRIEF The lurbinectedin JCA has the framework's ideal structure: one randomised trial, one population, one question. That makes its tensions purely interpretive. The assessment turns on the choice of data cut — significant at the earlier cut, not at the later one the assessors prefer — and never grades the certainty around that borderline signal. The cleanest case exposes the same central gap as the hardest ones.

While the tarlatamab and orphan assessments tested the framework where evidence is fragmented or scarce, lurbinectedin tests it where evidence is clear. The dossier rests on a single randomised trial, IMforte, comparing lurbinectedin plus atezolizumab with atezolizumab monotherapy as first-line maintenance. There is one population and one PICO. Nothing about the evidence structure is difficult.

Yet the assessment is not straightforward, because the trial supports two readings depending on which data cut is treated as authoritative. At the earlier, multiplicity-controlled cut, overall survival is statistically significant. At the mature cut, the one the assessors designate as their main source, it is not. The report documents this tension but, consistent with the framework, never resolves it into a graded statement of how much confidence the survival benefit deserves. The decisive act is interpretive and left implicit. Around it sit the same recurring tensions — selective exclusion of weak evidence, quality-of-life signals that run against the therapy, and stakeholder input recorded

but untraced. The strategic lesson is that the framework's central gap is not a symptom of hard evidence; it appears even when the evidence could hardly be clearer.

SECTION 01

Lurbinectedin as the Clean-Structure Case

> **IN BRIEF** A single randomised trial, a single population and a single PICO give the JCA its intended structure. Whatever tensions arise here cannot be attributed to scarce, fragmented or indirect evidence.

The Regulation's comparative machinery was designed for exactly this situation: a head-to-head randomised trial addressing a single well-defined question. Lurbinectedin provides it. The maintenance setting — patients whose disease has not progressed after first-line induction — yields a clean randomised comparison of adding lurbinectedin to atezolizumab versus atezolizumab alone, with overall and progression-free survival as co-primary endpoints.

This makes the case diagnostically valuable in the same way an easy case always is. If the assessment nonetheless leaves its most consequential judgement unstated, that cannot be blamed on the absence of a comparator, the fragility of an indirect comparison, or the multiplication of PICOs. It must be a property of the framework's interpretive design. Lurbinectedin therefore isolates, more cleanly than any fragmented case could, the question of what the JCA does when a single good trial yields a genuinely borderline result.

SECTION 02

One Trial, One Question — and One Decisive Choice

> **IN BRIEF** The whole assessment pivots on the choice of data cut. Survival is significant at the earlier cut; non-significant at the later one. The framework's cleanest case reduces to a single interpretive decision that is never graded.

The developer built its dossier primarily on the earlier data cut, at which overall survival was statistically significant under multiplicity control, with a hazard ratio of 0.73, and progression-free survival strongly favoured the combination. The assessors instead designated the later, more mature data cut as the main data source for the assessment, on the reasonable ground that its longer observation period yields greater data maturity. At that cut, overall survival is no longer nominally significant — a hazard ratio of 0.81 with a nominal p-value of 0.0605 — and, because formal hypothesis testing was not carried out on the later cut, all its p-values are nominal and uncontrolled for multiplicity.

So the framework's cleanest structural case reduces, in substance, to a single interpretive decision: which cut to believe. That decision is defensible either way — maturity is a genuine virtue, and so is multiplicity control — but it is decisive, and the report does not surround it with a graded certainty statement that would tell a national body how much weight the resulting estimate can bear. The strategic point is stark: even here, the assessment's conclusion is governed less by the evidence than by an interpretive choice about it, and that choice is presented as a methodological preference rather than as the pivotal judgement it is.

SECTION 03

The Maturity Principle as Implicit Judgement

IN BRIEF Preferring the more mature data cut is sound science — but here it systematically selects the less favourable estimate, and the exposure to non-protocol therapy it introduces is never weighed against the precision it buys. Whether that is neutrality or a conservatism reflex is exactly the kind of judgement the framework leaves unexamined.

Preferring more mature data is a defensible principle: longer follow-up reduces the risk that an early, immature estimate overstates a benefit. But in this case the principle is not neutral in its effect. The mature cut is precisely the point at which the survival benefit loses nominal significance, so applying the maturity principle systematically selects the less favourable interpretation of the same trial. What the additional follow-up bought can be stated precisely: deaths rose from 249 to 328, median follow-up rose from about 15 months to about 21 months, and the confidence interval around the hazard ratio narrowed by roughly 14% on the log scale. What it cost can be stated equally precisely: the analysis forfeited its confirmatory status, and the trial accrued a further six and a half months of exposure to non-protocol subsequent therapy. Between the two dates, 79 more patients died, but only 37 more reached a first progression event, so most of the additional survival information came from patients who had already progressed, which is precisely the window in which subsequent therapy can act on survival. The report records the gain and not the loss.

This is not a criticism of the choice, which is scientifically reasonable. It is an observation that the choice is a judgement with a predictable direction, and that the framework treats it as a technical preference rather than as a substantive act requiring justification and grading. The conservatism reflex seen in other cases — where, faced with ambiguity, the framework leans towards the more cautious reading — reappears here in a new guise: not the exclusion of weak evidence, but the selection among data cuts. A framework that cannot explicitly say, “we have chosen the mature cut, this is the confidence the resulting estimate deserves, and here is how the choice would change under the alternative” leaves the reader unable to see that a pivotal, direction-bearing judgement has been made at all.

Table. The two IMforte data cuts are compared. Values are as reported in the JCA report, the developer’s dossier, or the EMA assessment report; the final column also notes the net change. Each cell gives lurbinectedin + atezolizumab versus atezolizumab.

Element	First data cut — 29 Jul 2024	Second (mature) data cut — 12 Feb 2025
Overall-survival result	HR 0.73 (95% CI 0.57–0.95); p=0.0174	HR 0.81 (95% CI 0.65–1.01); nominal p=0.0605 — each estimate lies within the other's confidence interval
STATUS OF THE ANALYSIS		
Place in the testing plan	Planned interim analysis; boundary met; confirmatory	Prespecified as a data cut, but outside the sequential testing strategy — no alpha remained
Statistical role	Hypothesis-testing (confirmatory)	Hypothesis-generating (exploratory)
Prespecification	Prespecified analysis	Not prespecified for testing (outside the sequential strategy)
Multiplicity	Multiplicity-adjusted (alpha-controlled)	Not multiplicity-adjusted (nominal p only)
Designated by the report	Not the main data source	“The main data source” for the assessment
Used by the developer as	The basis of its own case	Exploratory and descriptive

OVERALL SURVIVAL IN DETAIL		
Deaths	113 (46.7%) vs 136 (56.4%) — 249 in total	159 (65.7%) vs 169 (70.1%) — 328 in total (+79)
Median overall survival	13.24 (11.89; 16.36) vs 10.65 (9.50; 12.16) months	13.90 (12.26; 16.13) vs 11.14 (10.12; 13.01) months — both arms rose
Difference in medians	2.60 months (0.35; 5.98); p=0.0199	2.76 months (0.48; 5.32); p=0.0269 — the difference increased
Statistical significance	Significant	Not significant
Boundary passed	The amended Hwang–Shih–DeCani boundary of 0.0313, and the O'Brien–Fleming boundary of 0.0214 it replaced	No boundary assigned
Precision (95% CI width, log scale)	0.511	0.441 — about 14% narrower; the whole of what maturity bought
THE STOPPING-RULE AMENDMENT		
Assessors' concern about the amendment	Applies to this cut's boundary — the assessors' stated concern, given neither direction nor magnitude	Not applicable (no alpha spent at this cut)
Regulatory finding on the same amendment	EMA: the changes were "operational and not data-driven"; the tipping-point analysis held at γ = −4.5, slightly stricter than O'Brien–Fleming; the change "did not affect the final study results"	Not applicable
Switching the amendment was meant to pre-empt	About 25% of the control arm expected — a figure the regulator itself calls an unexplained approximation	9.1% observed, at both cuts
EXPOSURE TO SUBSEQUENT THERAPY		
Control-arm access to lurbinectedin	Already present — all 22 switches (9.1%) were complete by this date; 44.6% vs 54.8% had received any subsequent anti-cancer therapy	The same 22 switches (9.1%); the count remains unchanged. What deepens is accrued exposure, not crossover
Accrued exposure to non-protocol therapy	About 15 months of median follow-up were accrued	A further ~6.5 months accrued on the same 22 switchers — the operative difference between the two cuts
Nature of subsequent therapy	Physician choice, non-protocol (NPT); not randomised	Physician choice, non-protocol (NPT); not randomised
Any subsequent anti-cancer therapy	108 (44.6%) vs 132 (54.8%)	130 (53.7%) vs 146 (60.6%) — +9.1 / +5.8 percentage points
First subsequent systemic therapy	Not reported separately	39.3% vs 48.1% — imbalanced towards the control arm
Median follow-up for survival	14.78 vs 15.18 months	21.59 vs 19.81 months — about +6 / +5 months
Patients progressed/died	384 (79.5%) / 249 (51.6%)	421 (87.2%) / 328 (67.9%) — only +37 progressions against +79 deaths
ADJUSTMENT FOR SUBSEQUENT THERAPY		
NPT-discounting adjustment (not switching-specific)	Provided — benefit strengthened: HR 0.71–0.72; p=0.009–0.010. A fixed discount was applied to any subsequent therapy in both arms, so it bounds the dilution rather than measuring it	Provided — rejected by the assessors as "not considered relevant", with no discussion
Switching-specific adjustment (RPSFT / IPCW)	Not performed; not requested	Not performed; not requested
Restricted mean survival time	Not provided	Not provided or requested, although the report recommends it elsewhere
PROGRESSION-FREE SURVIVAL		

Blinded independent review (IRF)	174 (71.9%) vs 202 (83.8%); HR 0.54 (0.44–0.67); p<0.0001	Not analysed at this cut
Investigator-assessed	178 (73.6%) vs 206 (85.5%); HR 0.55 (0.45–0.68); p<0.0001 — concordant with the IRF reading	204 (84.3%) vs 217 (90.0%); HR 0.56 (0.46–0.69); p<0.0001
Proportional hazards (OS)	Tested; not violated for OS	Tested; not violated for OS
RISK OF BIAS		
Overall survival — assessors	**Unclear — "other bias", the amendment**	**Low — reversing the methodological order**
PFS (IRF) — assessors	**Low**	**Not rated (not analysed at this cut)**
PFS (investigator) — assessors	**Low (high for blinding, settled by agreement with the blinded reading)**	**Low — but resting on the concordance observed at the July 2024 cut, the very cut the report declines to use**
Overall survival — this review	**Low**	**Very high (inferential status forfeited; greatest accrued exposure to non-protocol therapy)**

The comparison makes the inversion explicit. On every axis that bears on reliability, the first (interim) cut is the stronger analysis — prespecified, multiplicity-controlled, significant, and free of the additional exposure to non-protocol therapy that accrued afterwards — and its crossover-adjusted version is stronger still (HR 0.71–0.72), yet the assessors rate it "unclear" and set its adjustment aside. The second (mature) cut is exploratory, not multiplicity-controlled, non-significant, and more heavily exposed to non-random subsequent therapy — not because more control patients crossed over, since the 22 lurbinectedin switches (9.1%) were already complete at the first cut, but because those patients accrued a further six and a half months of survival time under it. Yet it is rated "low" and made the main data source. Proportional hazards were tested and not violated for overall survival in either case. The assessors' risk-of-bias direction is thus the reverse of the methodological reality.

SECTION 04

The Alpha-Spending Dispute: Where Statistics Becomes Governance

> **IN BRIEF** The trial changed its statistical stopping rule to a less stringent one shortly before recruitment closed. The assessors rightly flag that this could allow significance to be declared on limited data — a methodological point that is, at root, also a governance issue about trial integrity. The regulator, however, examined the same amendment and found the changes operational rather than data-driven, and that the interim would also have passed the boundary it replaced — findings the report does not cite.

The assessors devote particular attention to a late amendment to the trial's alpha-spending approach — a change, shortly before the completion of recruitment, from a more stringent stopping boundary to a less stringent one. Their concern, correctly stated, is that such a change may have been influenced by emerging trial information, and that a less stringent boundary allows statistical significance to be declared earlier, on comparatively limited event data. For this reason, the overall risk of bias for survival at the earlier, significant cut is rated "unclear."

The published regulatory record directly addresses that concern, and the report does not engage with it. Assessing the same amendment, the European Medicines Agency concluded that all changes were operational rather than data-driven and that they did not affect the final study results. The regulator also ran a tipping-point analysis showing that significance at the interim held even at $\gamma = -4.5$, a boundary it describes as slightly stricter than O'Brien–Fleming. The arithmetic points the same way: the interim p-value of 0.0174 passed not only the amended Hwang–Shih–DeCani boundary of 0.0313 but also the O'Brien–Fleming boundary of 0.0214 that the amendment had replaced. The regulator did criticise the stated rationale for the amendment, noting that using a stopping threshold to address treatment switching is not methodologically standard, as switching is ordinarily handled through sensitivity analyses; and the switching it was meant to guard against proved rarer than assumed, at 9.1% of the control arm against an expected figure of about 25%

that the regulator itself calls an unexplained approximation. None of this appears in the assessment report, which neither cites the finding nor explains why the possibility of a data-driven amendment remains open despite it.

The criticism was legitimate to raise, and it is where clinical assessment shades into a governance judgement about the conduct of the trial: the assessors are not only evaluating the developer's analysis but also adjudicating its statistical design choices against their own philosophy of inference. That is within their remit. The strategic difficulty is twofold. First, the framework gives such adjudications great weight, since an "unclear" rating on the pivotal endpoint effectively neutralises the developer's strongest, pre-specified result, while offering no graded vocabulary for expressing how much the concern should reduce confidence, as opposed to setting the result aside. Second, the same report treats the later cut, which is not pre-specified, is unadjusted for multiplicity, and was chosen after the data were seen, as low risk of bias; the pre-specified analysis is thereby held to a stricter standard than the post hoc one. The asymmetry is sharper once the regulatory finding is taken into account: the objection to the earlier cut has been tested and answered on the public record, whereas the objection that bears on the later cut, namely the additional exposure to non-protocol therapy accrued between the two dates, has never been tested at all. A defensible assessment would pre-specify both the authoritative data cut and the inferential framework before the data are available, and would grade the residual concern rather than switch cuts.

SECTION 05

Uncertainty Without Certainty, Around a Borderline Signal

> **IN BRIEF** The report catalogues the factors affecting certainty but never grades them. Around a survival benefit that is significant at one cut-off and not at another, the absence of a graded conclusion is not a detail — it is the whole question left unanswered.

The report’s treatment of uncertainty follows the now-familiar pattern: for each outcome, it lists the factors affecting certainty — the exclusion of patients with poor performance status and brain metastases, the open-label design, the shift to nominal testing at the mature cut, and unequal observation windows — but never translates them into a graded conclusion. There is no starting certainty, no transparent movement between levels, and no final rating.

In the fragmented cases, the missing certainty step mattered because it left many weak estimates unranked. Here, it matters for a sharper reason, and the way the choice is framed is itself part of the problem. The report implicitly sets two readings side by side as if equally weighted, a probable benefit attenuating with maturity or an early signal that did not hold up, but the numbers do not support that symmetry. Between the two cuts, the hazard ratio moves only from 0.73 (95% CI 0.57 to 0.95) to 0.81 (95% CI 0.65 to 1.01); the estimated effect and its upper bound barely change, and what crosses a threshold is the p-value (0.0174 to 0.0605), not the effect. A near-identical hazard ratio at a later cut with more events is far more consistent with a sustained benefit than with a signal that evaporated, which is exactly why the Kaplan-Meier curves should be examined directly rather than reduced to a significance verdict. The report supplies every input to this judgement and withholds the judgement itself, displacing the decisive interpretive step onto the reader and doing so in a way that allows a threshold-crossing p-value to masquerade as a change in the evidence.

SECTION 06

Evidence-Based Medicine and the Excluded Toxicity Data

> **IN BRIEF** The immune-related and patient-reported toxicity analyses were excluded due to methodological flaws rather than retained and graded — removing from view the tolerability evidence most relevant to maintenance therapy.

Because the evidence is a randomised trial, EBM's preference for randomised comparison is met. The selective application of EBM appears, as elsewhere, at the point of exclusion. The immune-related adverse-event analysis was excluded because the developer's post hoc selection was not considered comprehensive enough to capture all potentially immune-related events, omitting several toxicities listed for the immunotherapy backbone. The patient-reported toxicity results (PRO-CTCAE) were excluded as not interpretable, given missing baseline data and the absence of a time-to-event analysis accounting for unequal observation.

Both exclusions are defensible on their own terms. However, EBM would typically retain such analyses at low certainty and bound their interpretation rather than remove them. The consequence is specific to this indication: lurbinectedin is added to an immunotherapy given as maintenance to patients who, by definition, are doing well on prior treatment, so the tolerability of the addition is central to whether the trade-off is worthwhile. Removing the immune-related and patient-reported toxicity analyses leaves that trade-off harder to see at exactly the point where it is most decision-relevant.

SECTION 07

The Quality-of-Life and Safety Trade-off

> **IN BRIEF** Among the analyses retained, quality-of-life measures deteriorated more quickly with the combination, and safety was clearly worse. A maintenance therapy with a borderline survival signal and a worse tolerability profile is precisely where an explicit benefit–harm framing is most needed — and least present.

The analyses that were retained point in a consistent direction. Time-to-deterioration analyses show statistically significant faster deterioration for the combination on physical functioning and role functioning, among other scales, favouring the comparator; the mixed-model analyses to the sixth cycle show no significant differences, and the generic health-status measure shows none. On safety, the combination is worse across the board: more adverse events overall, more severe and serious events, more treatment discontinuations and interruptions, and a higher rate of fatal events, with interpretation limited by unequal observation duration.

This is the setting in which an explicit benefit–harm framing matters most. A maintenance therapy asks patients who are currently stable to accept added toxicity in exchange for a survival benefit whose magnitude and reliability are contested between the two data cuts. The report presents the survival, quality-of-life and safety strands separately and does not — because the framework forbids it — integrate them into an overall judgement. That prohibition is deliberate and appropriate at the assessment stage. But without a graded certainty statement for each strand, the national body must both integrate the strands and grade them, and it must do so for a therapy whose quality-of-life and safety signals run counter to a borderline survival benefit.

SECTION 08

The Participation Paradox

> **IN BRIEF** A carer independently identified the survival attenuation at the mature cut and the unfavourable quality-of-life signals, and suggested defining the PICO before the trial is designed. The input is annexed; its influence on the report is not traced.

One carer and one clinical expert provided written input on the scope and the revised draft. The carer's contribution was notably acute: it observed that overall survival at the second cut was no longer favourable, that most quality-of-life scales looked worse than the comparator, that the patient-reported toxicity baseline was missing, and — strikingly — suggested that the PICO should be defined before the trial is designed. The clinical expert situated lurbinectedin in practice, noting its more usual later-line use and cautioning that maintenance may not suit frailer patients.

As in the other cases, the input is annexed verbatim and stated to have been considered, but the report contains no trail showing how it shaped the scope, outcomes or interpretation. The paradox is the same, and here it carries particular sting: a carer identified, unprompted, several of the very tensions this review has described, including the data-cut attenuation and the quality-of-life trade-off. That a lay reviewer independently surfaced the assessment's central difficulties, while the report cannot show whether those observations changed anything, is the clearest possible illustration of participation recorded but influence untraced.

SECTION 09

Harmonisation and Fitness for Purpose

> **IN BRIEF** For a single clean RCT, the joint assessment delivers real harmonisation value — but by stopping at documented uncertainty around a borderline signal, it again leaves the decisive weighting to each national body.

Of all the early cases, this is the one where harmonisation should deliver most: a single randomised trial, assessed once to a high standard, sparing every national body the duplication of re-analysis. In large part, it does. The risk-of-bias analysis, the scrutiny of the stopping-rule amendment and the identification of the data-cut tension are exactly the kind of rigorous, reusable work the framework promises..

The limit is the same seam seen elsewhere. By stopping at documented uncertainty rather than a graded certainty conclusion, the joint layer leaves the decisive weighting of a borderline survival benefit to each national body. On the framework's own stated aim — a transparent assessment of relative effectiveness — the report succeeds. For a national body seeking a usable measure of how much lurbinectedin adds, it provides a rigorous account of why the answer is uncertain rather than a statement of how uncertain it is. A further harmonisation problem lies in the choice of cut. Under a treatment-policy estimand, the subsequent therapy patients receive is part of the question being answered, so the answer depends on what that therapy actually was. The great majority of the 483 randomised patients were enrolled in Europe and the Middle East, and only 31 (6.4%) in North America ⟨regional counts to be re-checked against the JCA regional table before publication⟩. Lurbinectedin was not authorised in the EU for second-line use when the assessment was made, so European control-arm patients obtained it mainly through compassionate use. The later care built into the mature estimate may therefore not correspond to routine practice in the Member States that must apply the result, and its distribution across regions is not reported, although regional analyses exist. For patients, it is largely silent on the integrated benefit–harm

question the carer raised. These are three audiences a single document cannot equally serve, and recognising that is the strategic insight.

SECTION 10

What Reform Would Actually Require

> **IN BRIEF** The reform agenda is narrow: pre-specify the authoritative data cut, grade certainty explicitly, retain and bound tolerability analyses, and trace stakeholder influence. None requires new methodology — only that a few decisive judgements be made openly.

The lurbinectedin case points to a compact reform agenda. First, both the authoritative data cut and the inferential framework for HTA should be pre-specified, ideally through Joint Scientific Consultation, so that neither the governing cut nor the standard of inference is chosen after the results are known; otherwise the assessment is, on its own terms, post hoc and hypothesis-generating rather than confirmatory. Second, and most importantly, documented uncertainty should be converted into an explicit degree of certainty; a borderline survival signal is the paradigm case in which this step is indispensable, and its absence here is the report's central gap. Critically, the same evidentiary standard must apply to both data cuts, so that a pre-specified, multiplicity-controlled analysis is not held to a stricter test than a later, non-prespecified one.

Third, tolerability analyses — immune-related adverse events and patient-reported toxicity — should be retained at low certainty and bounded rather than excluded, because they are decision-critical for maintenance therapy. Fourth, a short, honest record should trace how stakeholder input shaped, or did not shape, the assessment, so that a carer who identifies the assessment's central tensions can be shown to have been heard in substance, not merely in form. None of these requires new analytical sophistication. Each requires the framework to commit, on the record, to a judgement it currently leaves implicit — a governance task rather than a scientific one.

IN CLOSING

Conclusion

> **IN BRIEF** The framework's cleanest structural case still leaves its most important judgement — how much confidence a borderline survival benefit deserves — unstated. From the opposite end of the evidence spectrum to the single-arm cases, lurbinectedin confirms that the missing uncertainty-to-certainty step is structural.

The lurbinectedin JCA is the mirror image of the single-arm orphan cases. Whereas those tested the framework with the weakest possible evidence, this tests it with close to the strongest structural form: one randomised trial, one population, one question. The assessment is rigorous, and the assessors' conduct exemplary. Yet the case reduces to a single interpretive choice — which data cut to believe — that determines whether the therapy's survival benefit reads as significant or equivocal. The report never grades the confidence that choice leaves the reader with.

The decisive finding is that the framework's cleanest case exhibits the same central gap as its hardest ones. The missing step from uncertainty to certainty, the exclusion rather than grading of weak evidence, and the recording of stakeholder input without tracing its influence are not artefacts of scarce or fragmented evidence. They are properties of the framework's interpretive design, visible whether the evidence is a single-arm study, a fragmented multi-PICO scope, or — as here

— a single clean randomised trial. Lurbinectedin's lasting contribution is to show that closing a small number of boundaries, not adding methodology, is what the next phase of European HTA requires.

TECHNICAL REVIEW

Lurbinectedin (Zepzelca®)

Joint Clinical Assessment

Claim-by-claim, against the binding methodology and the source record

Case	JCA-MP-2024-16 — Lurbinectedin (Zepzelca®), Pharma Mar S.A.; first-line maintenance of extensive-stage small cell lung cancer
Assessor / Co-assessor	IQWiG, Germany (B. Wieseler) / INFARMED, Portugal (R. Moura)
Legal & methodological basis	Regulation (EU) 2021/2282 (HTAR); Implementing Regulation (EU) 2024/1381; HTACG guidance under Art. 3(7)(d)
Procedural status	Report v1.0; endorsed by the HTACG on 22 June 2026; EC procedural review 1 July 2026
Register	Technical, strong factual claims; the conceptual/governance argument is held in the companion strategic review

BOTTOM-LINE SUMMARY

Bottom-Line Summary

BOTTOM LINE The assessment team's conduct is rigorous and exemplary; the substantive tension lies in the evidence itself. The report is procedurally compliant and methodologically defensible. A single randomised trial (IMforte) directly informs the single PICO; the assessment hinges on choosing the mature data cut (OS HR 0.81, nominal p=0.0605) rather than the developer's earlier multiplicity-controlled cut (OS HR 0.73, p=0.0174), with no graded certainty conclusion regarding a borderline signal.

SECTION 01

Purpose, Scope and Method of This Review

KEY MESSAGES

Two actors, judged separately. The assessment team (IQWiG/INFARMED, I.P.) versus the developer's dossier.

One benchmark. Every claim anchored to the source record and measured against HTACG guidance under Art. 3(7)(d) HTAR.

Headline verdict. Procedurally compliant and methodologically defensible; the analytical tension is data-cut selection, not evidence structure.

This review evaluates the lurbinectedin JCA against the rules that bind it, claim by claim, and against the source record: the JCA report, the developer's dossier and appendices, and the request-and-response documentation. While the companion strategic review asks what the case reveals about the framework, this document asks what exactly was done, under which provision, with which figures, and with what documented effect.

Central thesis. The assessment team conducted the assessment in compliance with — and by actively enforcing — the binding methodology. In contrast to the fragmented cases, the evidence structure here is ideal: a single randomised trial directly informs the single PICO. The substantive tension is therefore not structural but interpretive, and it centres on the choice of authoritative data cut, the scrutiny of the trial's stopping rule, and the exclusion of two toxicity outcomes.

SECTION 02

Procedural and Structural Compliance

KEY MESSAGES

All testable procedural requirements are met. Template, single-PICO scope, stakeholder involvement and transparency each check out.

No procedural non-compliance. The compliance question turns on the substantive methodology.

Requirement	Finding
Template	Follows the Annex template of Implementing Regulation (EU) 2024/1381 — general information, background, scope, information retrieval, study characteristics and risk of bias, results.
Scope	Set under Article 8(6) HTAR: one population (full claimed maintenance indication) and one PICO — lurbinectedin + atezolizumab versus atezolizumab monotherapy.
Stakeholder involvement	Documented under Articles 8(6) and 11(4): one carer and one clinical expert provided input on the scope and the revised draft reports.
Joint Scientific Consultation	None held; consequently, no deviations from JSC propositions to report.
Risk of bias	RoB 1 applied at the outcome level within a treatment-policy framework, separately for each of the two data cuts.

SECTION 03

The Evidence Base: One RCT, Two Data Cuts

KEY MESSAGES

One pivotal randomised trial. IMforte (GO43104): randomised, open-label, phase III; lurbinectedin + atezolizumab (n=242) vs atezolizumab (n=241) maintenance.

Two data cuts. Earlier (multiplicity-controlled, OS significant) and mature (assessors' main source, OS not nominally significant).

Direct evidence only. No single-arm data or indirect comparisons.

The assessment rests on a single pivotal study, IMforte, a randomised, open-label, phase III maintenance trial in which patients whose disease had not progressed after first-line induction were randomised to lurbinectedin plus atezolizumab (242 patients) or to atezolizumab monotherapy (241 patients). The trial defines two analysis time points: an earlier cut, at which the co-primary endpoints were analysed under multiplicity control, and a later, more mature cut. The assessors designate the mature cut as the main data source for the assessment on the grounds of greater data maturity, while noting that at the mature cut formal hypothesis testing was not performed, so its p-values are nominal and uncontrolled for multiplicity.

There is no single-arm evidence, no matching-adjusted comparison, and no network. The single PICO is informed solely by direct head-to-head randomised evidence, the structure the framework was designed for, and the reason every substantive tension in the case is interpretive rather than structural.

SECTION 04

The Single PICO and Its Results

KEY MESSAGES

Survival benefit is cut-dependent. Significant at the earlier cut; not nominally significant at the mature cut.

Response outcomes show no difference; quality of life and safety favour the comparator. The combination shows a consistent disadvantage in tolerability.

Two outcomes were excluded. Immune-related AEs and PRO-CTCAE were removed for methodological reasons.

The results for the single PICO, as reported:

Outcome	Result (by data cut where relevant)
Overall survival — earlier cut	HR 0.73 [0.57, 0.95], p=0.0174 — statistically significant (multiplicity-controlled).
Overall survival — mature cut (main source)	HR 0.81 [0.65, 1.01], nominal p=0.0605 — not statistically significant.
Progression-free survival (IRF, earlier cut)	HR 0.54 [0.44, 0.67], p<0.0001; investigator-assessed 0.55 [0.45, 0.68] at the same cut. At the mature cut, investigator-assessed only, HR 0.56 [0.46, 0.69]; no IRF analysis was performed.
ORR/complete response/duration of response	No statistically significant differences; duration of response HR 1.34 [0.67, 2.69] (numerically favours comparator; non-randomised responder subset).
Disease-specific HRQoL (time-to-deterioration)	Faster deterioration on the combination: physical functioning HR 1.583 [1.106, 2.264]; role functioning HR 1.406 [1.032, 1.915]; MMRM to cycle 6: no significant differences.
Health status (EQ-5D-5L)	No statistically significant differences.
Safety (descriptive, mature cut)	Worse for the combination: any AE 97.1% vs 81.7%; severe (≥G3) 46.3% vs 27.1%; serious 34.3% vs 17.9%; fatal (G5) 5.0% vs 2.5%; discontinuation 7.0% vs 4.2%; interruption 41.3% vs 15.8%.
Immune-related AEs; PRO-CTCAE	Excluded from the report (methodological flaws/not interpretable).
PFS2; generic HRQoL	Not recorded in the trial; not assessed.

SECTION 05

The Data-Cut and Alpha-Spending Disputes

KEY MESSAGES

Data cut is the pivotal disagreement. Developer relied on the earlier significant cut; assessors adopted the mature, non-significant one.

The stopping rule was changed late. A shift to a less stringent alpha-spending function shortly before recruitment closed; assessors rated the earlier-cut OS as “unclear” risk of bias.

The central methodological dispute concerns which data cut is authoritative. The developer built its dossier primarily on the earlier cut, where overall survival was significant, and argued that the later cut was exploratory. The assessors adopted the mature cut as the main data source, citing

its greater maturity, and treated its non-significant survival result accordingly. The choice is decisive and is not accompanied by a graded certainty statement reconciling the two readings.

The report fails to justify the asymmetry in how the two data cuts are treated. The concern about the alpha-spending amendment rests on the premise that information emerging during the trial could have influenced the change. That premise has been examined elsewhere and does not survive scrutiny: assessing the same amendment, the European regulator concluded that all these changes were operational rather than data-driven, and that the change did not affect the final study results. The regulator further reported a tipping-point analysis in which significance at the interim held even at γ = −4.5, a boundary it describes as slightly stricter than O'Brien–Fleming; and the interim p-value of 0.0174 in fact passed both the amended Hwang–Shih–DeCani boundary of 0.0313 and the O'Brien–Fleming boundary of 0.0214 that the amendment replaced. The assessment report cites none of this and does not explain why the possibility remains open despite it. The objection that decided the choice of data cut was recorded but never measured, and the one body that did measure it reached the opposite conclusion. Nor is the scepticism symmetrical: assessors with full access to the trial data are at least as able to anticipate how a later cut will read, so preferring that later cut is not self-evidently more neutral than the developer's pre-planned, multiplicity-controlled analysis. Treating a post hoc cut chosen after full data access as carrying only mild bias, while rating the pre-specified, alpha-controlled analysis as "unclear" on the strength of a modest boundary amendment, applies scepticism unevenly. The amendment concern is legitimate and worth raising; but a rating that neutralises the developer's strongest result should rest on a detailed, even-handed argument, not a brief one that leaves the regulatory finding to the contrary unmentioned. The framework offers no graded vocabulary to express how far the concern should reduce confidence, rather than set the result aside. A further point tells against treating the mature cut as a reversal at all. The two survival estimates are close, and both indicate benefit: the hazard ratio moves only from 0.73 (95% CI 0.57 to 0.95) to 0.81 (95% CI 0.65 to 1.01), and the upper confidence bound shifts only from 0.95 to 1.01, while what changes decisively is the p-value, from 0.0174 to 0.0605, as it crosses the 0.05 threshold. Treating that crossing as a qualitative change from benefit to no benefit is the dichotomisation the statistical literature has long cautioned against: a modestly attenuated but still clinically relevant point estimate, whose confidence interval only just includes unity, is evidence of a probable benefit, not of its absence. It is also why the Kaplan-Meier curves matter and should be inspected directly: near-identical hazard ratios across two data cuts, with the later one having more events, point to a sustained separation of the survival curves rather than to a signal that failed to hold.

What Changed Between the Two Data Cuts, and Why

The two overall-survival analyses are separated by only about six and a half months: the interim data cut on 29 July 2024 and the final, last prespecified cut on 12 February 2025. The interim analysis was the formal, alpha-controlled test of overall survival, performed at the planned trigger of approximately 219 deaths and read at 249 actual deaths under the trial's group-sequential design. It crossed the (amended) stopping boundary — and, at p=0.0174, would also have crossed the O'Brien–Fleming boundary of 0.0214 that the amendment replaced. The 12 February 2025 analysis was the last prespecified data cut, but because the alpha had been spent at the interim, it fell outside the sequential testing strategy, which is why its p-value is only nominal. Both analyses used the same estimators: a stratified Cox model for the hazard ratio, a stratified log-rank test, and Kaplan-Meier medians, under a treatment-policy estimand.

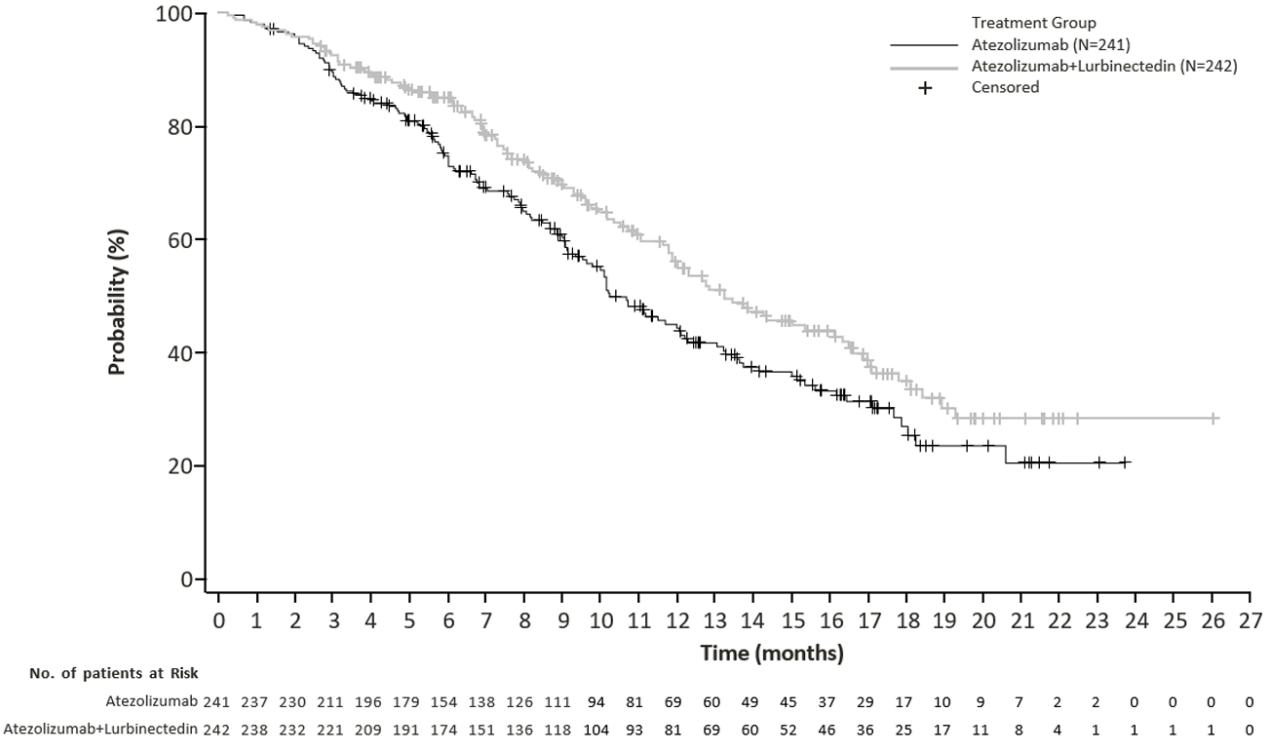
Three explanations can account for a hazard ratio moving from 0.73 to 0.81 while both medians are maintained. The first is ordinary sampling variation: the interim estimate rests on fewer events and is less precise, so some drift towards a slightly less extreme value as events accrue is expected and does not imply the early result was spurious. The second is non-proportional hazards: if the survival curves separate early and then run in parallel or converge slowly in the

tail, the cumulative hazard ratio drifts towards one as later deaths accumulate, even when the early separation persists. That is why the shape of the Kaplan-Meier curves and a restricted mean survival time (RMST) analysis are more informative here than a single hazard ratio, and why the report should present them. The third, and most concrete, is differential post-progression therapy.

On the third point, the trial data are telling, and the mechanism is specific. Because 22 control-arm patients (9.1%) received lurbinectedin itself, and a larger share of the control arm received systemic anti-cancer treatment overall (54.8% versus 44.6% at the first cut, rising to 60.6% versus 53.7% at the second), part of the control arm's survival reflects the benefit of the very agent under test. For those patients, the overall-survival comparison becomes, in effect, a partial comparison of lurbinectedin against lurbinectedin, which biases the estimate towards the null. Randomisation must not be broken to remove this, and the intention-to-treat analysis is correct precisely because it preserves randomisation; the sound inference is therefore that the intention-to-treat overall-survival hazard ratio is a conservative, diluted estimate of the treatment effect, not that the effect is smaller than the interim indicated. The reports do not disclose the individual outcomes of those 22 patients, nor how long any patient spent on subsequent therapy, so the dilution can be argued in direction but not quantified from the published tables. It is important to be precise about what changes between the two cuts. The number of control-arm patients who received lurbinectedin was 22 (9.1%) at both dates and did not change, so the earlier analysis was already affected. What grew over the intervening six and a half months was the accrued exposure, the person-time lived under non-protocol therapy, and that, rather than any additional crossover, is the mechanism by which the dilution deepens at the mature cut. Crucially, the one analysis that could have quantified it was set aside: for the mature data cut, the developer submitted a supplementary overall-survival analysis discounting the survival time accrued after non-protocol anti-cancer therapy, and the assessors dismissed it in a single sentence as "not considered relevant for the assessment." The assessment thus rests on the diluted intention-to-treat estimate while declining the very analysis designed to measure the dilution. Loss to follow-up, by contrast, cannot explain the change: it was minimal and balanced (one patient, 0.4%, in each arm), so censoring for overall survival is essentially administrative and informative censoring is not a plausible driver for this endpoint.

Two further considerations settle the interpretation. First, the magnitude at stake is trivial: the entire difference between a "significant" and a "non-significant" verdict here is a shift of less than one tenth on the hazard-ratio scale, with the point estimate moving from 0.73 to 0.81 and the upper confidence limit from 0.95 to 1.01. The estimated benefit, a reduction of roughly one fifth to one quarter in the hazard of death, is materially unchanged; only its position relative to an arbitrary threshold has shifted. The difference in median survival actually increased over the same interval, from 2.60 months (13.24 versus 10.65) to 2.76 months (13.90 versus 11.14). Each hazard-ratio estimate lies within the confidence interval of the other, so the two analyses are better read as one estimate observed twice than as two conflicting results. Second, the mature-cut result is, by the trial's own design, descriptive and hypothesis-generating rather than confirmatory: with the alpha already spent at the interim, no inferential weight attaches to its nominal p-value, so it can enrich but cannot overturn the confirmatory analysis that met its prespecified boundary. Whether longer follow-up would move the estimate again is constrained by the disease, since in extensive-stage small-cell lung cancer most deaths occur early. By February 2025, a large share of events had already accrued, and the hazard ratio is more likely to stabilise near 0.81 than to swing back. Taken together, the value and the benefit are maintained; what changed is a threshold-crossing p-value, compounded by a dilution bias that the correct intention-to-treat analysis cannot and should not remove. Adjudicating this requires the Kaplan-Meier curves and a restricted mean survival time estimate, not a significance verdict.

Figure 1 Kaplan-Meier plot of overall survival in IMforte (cut-off: 29 July 2024)



What to read from the curve: (i) the timing of separation between the two arms; (ii) whether the curves remain separated or converge in the tail; (iii) the number-at-risk row, which shows how many events underlie the tail estimate; and (iv) consistency between the 29 July 2024 and 12 February 2025 cuts. Together these distinguish a genuine, durable benefit from an early signal that regressed.

Proportional Hazards and the Uninformative Tail

A related technical point concerns the shape of the effect over time. At the final data cut, the Schoenfeld residual test was nominally significant ($p < 0.05$) for investigator-assessed progression-free survival and for the safety endpoints, indicating a departure from proportional hazards. The assessors nonetheless judged the hazard ratio interpretable, on the grounds that the Kaplan-Meier curves show no distinct crossover and that the estimator can be read as at least a weighted average effect over the observed period. That is a reasonable position, but it is the opposite of the approach applied to overall survival, where a small, threshold-crossing change in a single hazard ratio was allowed to drive the conclusion. If a weighted-average hazard ratio is acceptable under non-proportional hazards for progression-free survival, the same tolerance should apply to the survival estimate rather than a strict significance criterion.

The progression-free survival curves are also instructive about the tails. In the developer's own account, they separate consistently and cross only late, at approximately eighteen months, when only a small proportion of patients remain at risk. A late crossing, supported by a handful of patients, carries little information and should not be over-interpreted; it is precisely the region of a survival curve where estimates are least stable. The lesson for overall survival is the same: the reliable signal lies in the body of the curve, where separation is established and events are dense, not in a sparse tail governed by few patients. This is a further reason to read the Kaplan-Meier curves directly and to summarise them with a restricted mean survival time rather than a single, tail-sensitive hazard ratio.

The Adjusted Mortality Analysis, and the Inverted Risk-of-Bias Ratings

Two technical facts, both in the dossier, bear directly on how the survival result should be read. First, proportional hazards were tested: the developer supplied log-cumulative-hazard plots and Schoenfeld residual tests for the time-to-event endpoints. At the final data cut the Schoenfeld test was nominally significant only for investigator-assessed progression-free survival and for the safety endpoints; overall survival was not flagged as violating proportional hazards, so the Cox hazard ratio is a legitimate summary of the mortality effect, though the report reasonably notes that no restricted-mean-survival-time analysis was provided. Second, the hazard ratio was estimated with a stratified Cox model and the log-rank test, both of which weight each event time by the size of the risk set. Because the risk set is largest early and shrinks over time, early events dominate and the sparse tail contributes little; the model does not favour the tail. The survival curves separated early, with a clear separation by three months in the dossier, and the twelve-month survival rate favoured the combination (56.25% versus 44.10%, a statistically significant difference). The benefit is front-loaded, exactly the pattern a Cox hazard ratio captures well, so the modest attenuation from 0.73 to 0.81 with longer follow-up is not an artefact of a noisy tail.

What does change with follow-up is the amount of post-progression therapy folded into a treatment-policy estimand, and that therapy was neither balanced nor protocol-allocated. Subsequent treatment was non-protocol therapy chosen by the treating physician, not assigned by the study. The control arm progressed sooner (median investigator-assessed progression-free survival 2.73 versus 5.55 months) and discontinued treatment earlier and slightly more often (91.3% versus 88.8%), so more of its patients moved to further lines: 48.1% versus 39.3% received systemic anti-cancer therapy as the first subsequent treatment, and, critically, 22 control-arm patients (9.1%) received lurbinectedin itself, which no patient in the combination arm received afterwards. To the extent that control-arm patients were treated with the experimental agent, the intention-to-treat contrast becomes, in part, lurbinectedin against lurbinectedin, biasing the overall-survival comparison toward the null. This dilution increases as the trial matures, not because additional patients cross over: the 22 switches (9.1%) were already complete at the first cut, and the count is identical at both dates. It increases because each of those patients accrues more survival time on the experimental agent as follow-up lengthens. It is therefore largest at the very cut the assessors chose as authoritative, and it was already present, not absent, at the cut they set aside, since by 29 July 2024 some 54.8% of the control arm and 44.6% of the combination arm had already received further anti-cancer treatment.

The developer did not leave this unexamined. It pre-specified a supplementary overall-survival analysis that discounts survival time accrued after non-protocol therapy, intended to approximate the effect had that subsequent therapy not been available. It is an imperfect instrument, because the discount is applied to any subsequent therapy in either arm rather than to the lurbinectedin crossover specifically, so it bounds the dilution rather than measuring it. When the crossover imbalance is adjusted for in this way, the mortality benefit does not weaken; it strengthens. At the first data cut, the discounted analyses returned hazard ratios of 0.71 (p=0.0091) and 0.72 (p=0.0104) across a range of discounting assumptions, with point estimates below the unadjusted 0.73 and p-values comfortably under 0.05. The equivalent adjusted analysis for the mature cut was also provided. The assessors dismissed it in a single sentence, as "not considered relevant for the assessment," with no accompanying discussion. An analysis that directly addresses the one bias the report itself identifies, and that points, if anything, to a larger benefit, was set aside without engagement, while the unadjusted, diluted estimate was retained as the basis for the conclusion. Nor was a better instrument sought: no rank-preserving structural failure time analysis, no inverse-probability-of-censoring weighting, and no restricted mean survival time estimate was requested under Article 11(2), although the report itself notes the absence of the latter.

The risk-of-bias ratings complete the inversion. The confirmatory interim analysis, prespecified, multiplicity-controlled and statistically significant, is rated at "unclear" risk of bias on the strength of the alpha-spending amendment (Table 20). The mature analysis, exploratory, outside the

sequential testing strategy, not multiplicity-controlled, and most affected by the subsequent-therapy dilution, is rated at "low" risk of bias and made the main data source (Table 19). The "low" rating is narrowly defensible in its own terms, because it speaks only to detection bias: death is an objective event that an open-label design cannot distort. But it is silent on the bias that actually threatens the mature estimate, the non-random, physician-driven contamination of the control arm, so the label certifies as clean the analysis that is in fact the more compromised and flags as doubtful the analysis that is methodologically stronger. On the two axes that matter for a survival conclusion, inferential status and freedom from treatment-switching bias, the direction of the risk-of-bias labelling is the reverse of the underlying reality. Two further observations sharpen the point. In the Cochrane instrument, the report's use of "unclear" does not mean "moderate": it indicates that the available information was insufficient to choose between low and high, so the rating reflects what the assessors had before them rather than what the trial did. This makes the failure either to record the regulatory finding on the amendment or to request anything further about it the operative gap. And the low rating for investigator-assessed progression-free survival at the mature cut cannot rest on independent confirmation, because no blinded review facility analysis was produced for 12 February 2025; it rests instead on the concordance between blinded and investigator readings observed at 29 July 2024, the very cut the report declines to use as its main source.

SECTION 06

Excluded Outcomes: Immune-Related AEs and PRO-CTCAE

> **KEY MESSAGES**
>
> **Two decision-critical toxicity outcomes were excluded and not graded.** For maintenance therapy added to an immunotherapy, both are central to the benefit–harm trade-off.

The report excludes two toxicity outcomes rather than retaining them at low certainty. The immune-related adverse-event analysis is excluded because the developer's post hoc selection was deemed insufficiently comprehensive to capture all potentially immune-related events, omitting several toxicities listed for the immunotherapy backbone; the data are judged insufficient for assessment. The patient-reported toxicity results (PRO-CTCAE) are excluded as not interpretable, given the absence of baseline data, missingness in the branching logic, and the lack of a time-to-event analysis for unequal observation.

Each exclusion is defensible, but the technical consequence is that the two outcomes most directly relevant to the tolerability of adding lurbinectedin to atezolizumab in a maintenance setting are omitted from the comparative assessment, while the retained safety outcomes already show a clear disadvantage for the combination. An evidence-based-medicine approach would typically present these analyses with explicit caveats rather than remove them, precisely so that the benefit–harm trade-off remains visible. This is a deviation from EBM.

SECTION 07

Certainty Versus Uncertainty (Quantified)

> **KEY MESSAGES**
>
> **Uncertainty documented, never graded.** The recurring formula is "the certainty of evidence for this outcome is potentially affected by …," with no high/moderate/low/very-low output.
>
> **The gap bites hardest at a borderline signal.** A survival benefit significant at one cut but not the other cannot be adjudicated without a graded conclusion.

The template requires that the degree of certainty regarding relative effectiveness and safety be assessed. The report documents the factors affecting certainty per outcome but does not provide a graded conclusion and does not use GRADE. The contrast with a structured framework is the same as in the other cases:

Element	GRADE	Lurbinectedin JCA report
Starting certainty	Explicit, by study design	None stated
Domains	RoB, inconsistency, indirectness, imprecision, publication bias	RoB, generalisability, testing at the mature cut, observation windows — narrative
Movement between levels	Transparent downgrade/upgrade	No algorithm visible
Output	High/moderate /low/very low	No comparable final category
Traceability	Reader can follow the derivation	Reader sees why it is uncertain, not how it becomes a degree of certainty

Consequence. Because the central survival estimate is borderline, significant at one cut and non-significant at the other, the absence of a graded certainty conclusion is not presentational but operational. The national body must decide whether this is a probable benefit that attenuates with maturity or an early signal that did not hold, and the report provides all inputs to that judgement while withholding the judgement itself. A pre-specified, multiplicity-adjusted endpoint, and thus hypothesis testing, is always more robust than a post hoc analysis with a data cut decided after the study results have been made available and thoroughly scrutinised.

SECTION 08

Risk of Bias by Outcome and Data Cut

KEY MESSAGES

Objective endpoints rated low; subjective and post-randomisation endpoints rated high. Consistent with an open-label maintenance trial.

The alpha-spending change drives the one "unclear" rating. Survival at the earlier cut.

Outcome	Risk of bias
Overall survival — mature cut	Low (detection bias only; silent on accrued non-protocol therapy)
Overall survival — earlier cut	Unclear (late alpha-spending amendment; the regulator found the change operational rather than data-driven)
Progression-free survival (IRF/investigator)	Low (IRF–investigator concordance at the 29 Jul 2024 cut only; no IRF analysis at the mature cut)
ORR/complete response/duration of response	High (open-label; DoR from non-randomised responder subset)
Disease-specific HRQoL; health status	High (open-label; differential attrition)
Serious AE, severe AE, AE-related death	Low
Other safety outcomes	High (unequal observation time; open-label attribution)

The pattern is coherent and superficially defensible: objective, hard endpoints (survival at the mature cut, the serious and fatal safety categories, and progression-free survival, supported by independent review) are rated as low risk of bias, whereas subjective and post-randomisation outcomes are rated as high risk of bias, consistent with an open-label maintenance design. The difficulty lies in how the two survival analyses are treated. Applying the same principles consistently, the mature-cut analysis, conducted outside the sequential testing strategy,

unadjusted for multiplicity, and settled upon after the trial results were available and scrutinised, has at least as strong a claim to a cautionary rating as the earlier, pre-specified, multiplicity-controlled analysis, which instead receives the single "unclear" rating solely for the stopping-rule amendment. In evidence-based medicine, a pre-specified analysis under multiplicity control is ordinarily the more robust of the two; here, the ratings run in the opposite direction. The asymmetry is compounded by what each rating rests on. The “unclear” rating for the earlier cut rests on a single objection that the European regulator had already examined and rejected, finding the protocol changes operational rather than data-driven; the “low” rating for the mature cut rests on the objectivity of death, which is true but silent on the accrued exposure to non-protocol therapy, the mature estimate’s actual vulnerability. The report itself states that survival is not susceptible to detection bias because death is an objective event that cannot be influenced by knowledge of treatment allocation, yet survival is the one outcome that receives the cautionary rating. That asymmetry between the two data cuts is the report's most consequential departure from a consistent evidentiary standard.

SECTION 09

Expert and Carer Input: Documented but Not Traced

KEY MESSAGES

A lay reviewer surfaced the assessment’s central tensions. The carer flagged the survival attenuation at the mature cut and the unfavourable quality-of-life signals.

Influence is not traced. Input annexed and stated to have been considered; no record maps it to scope or interpretation.

Participant	Input and concerns raised
Carer	Questioned the exclusion of patients with poor performance status and brain metastases as unrepresentative of real-world use; asked that lower-grade adverse events also be reported for their relevance to quality of life; observed that overall survival at the mature cut-off was “not so favourable anymore”; noted that most quality-of-life scales were worse than those of the comparator; flagged the missing baseline patient-reported toxicity; and suggested defining the PICO before the trial is designed.
Clinical expert	Situated lurbinectedin in its more usual later-line use; noted the maintenance benefit suggested by the trial; cautioned that maintenance may not suit frailer patients; and noted higher myelosuppression in the combination arm, which was partly offset by growth-factor prophylaxis.

The input is annexed verbatim and stated to have been considered before finalisation. Several of the concerns it raises — the limits on generalisability, the quality-of-life disadvantage, the data-cut attenuation — are indeed reflected in the report’s uncertainty documentation. However, there is no change-tracking record mapping specific input to specific scope or interpretive decisions. The technical finding is the same as in the companion cases: participation is documented at the level of opportunity, not of demonstrable influence. It is not a procedural breach, but it is a transparency limitation — and it is especially conspicuous here, because a lay reviewer independently identified the assessment’s central difficulties.

SECTION 10

Additional-Information Handling and Legal Fragility

KEY MESSAGES

One request round, two items — fully addressed. First, subsequent therapy by arm; induction-phase disposition and a patient-count discrepancy.

Legal exposure is narrow. The strongest grounds challenge the statement of reasons for the data-cut choice; annulment would likely fail.

The assessors issued a single request for additional information with two items: the first subsequent anti-cancer therapy by arm, and the participant disposition and reasons for discontinuation during the induction phase, where a discrepancy in patient counts (raising a selection-bias concern about who entered the randomised maintenance phase) required clarification. The developer supplied the subsequent-therapy data and an updated disposition flowchart reconciling the counts, and the assessors recorded that the requested points were fully addressed. The request was issued on 25 February 2026 and answered on 12 March 2026. Neither item addressed the stopping-rule amendment. The power under Article 11(2) is discretionary, so no duty was breached by leaving that question open. However, the published record contains no request for information on the duration of subsequent therapy, the outcomes of the 22 control-arm patients who received lurbinectedin, any switching-specific adjustment method, or restricted mean survival time, nor any explanation of why no further clarification was needed for the single judgement on which the choice of data cut rests. At the factual-accuracy stage, the developer separately confirmed that certain quality-of-life subgroup analyses had been incorrect and submitted corrected versions, handled outside the report body, while confirming that the original interaction p-values were correct.

On legal robustness, the analysis mirrors the companion cases but with a narrower exposure. The strongest argument grounds would challenge the quality of reasoning, an inadequate statement of reasons under Article 296 TFEU for preferring the mature cut and for treating the stopping-rule amendment as decisive, and the inconsistent treatment of uncertainty in the absence of an explicit threshold. Weaker grounds would allege manifest error in the risk-of-bias ratings. A defence would likely prevail: EU bodies enjoy wide scientific discretion; review is limited to manifest error, misuse of powers and procedural irregularity; there is no right to a particular data cut or stopping rule; and the report may not be a challengeable binding act because it does not itself determine price, reimbursement or market access. The residual exposure is concentrated entirely in the reasoning around the data-cut choice and the certainty gap. This would require a deeper analysis, as several additional arguments deserve attention.

SECTION 11

Bottom-Line Technical Verdict

Dimension	Verdict	Basis
Procedural/structural (assessors)	Compliant	Template, single-PICO scope, stakeholder involvement, and transparency all met
Methodological conduct (assessors)	Compliant/exemplary	Independent RoB by data cut; defensible exclusions. But the stopping-rule concern remains unmeasured, and the regulatory finding that the amendment was operational rather than data-driven is neither cited nor rebutted
Evidence structure	Ideal	One randomised trial directly informs the single PICO
Outcome — survival	Borderline/cut-dependent	Significant at the earlier cut (HR 0.73); not nominally significant at the mature cut (HR 0.81). The median difference increased from 2.60 to 2.76 months; what changed was a threshold-crossing p-value and an additional six and a half months of exposure to non-protocol therapy
Outcome — QoL and safety	Favour comparator	Faster HRQoL deterioration with the combination; safety clearly worse
Certainty reporting	Gap	Uncertainty documented; no graded certainty conclusion and no weighing of the precision gained at the mature cut (~14% on the log scale) against the confirmatory status and internal validity forfeited
EBM alignment	Selective	Randomised comparison honoured; toxicity analyses excluded rather than graded
Expert/carer input	Documented, not traced	A lay reviewer surfaced the central tensions; influence not demonstrated
Legal robustness	Defensible	The strongest challenge targets the reasoning behind the data-cut choice

In sum. The lurbinectedin JCA is procedurally compliant and methodologically defensible, and the assessment team's conduct is rigorous and exemplary. Despite the framework's ideal evidence structure — one randomised trial, one PICO — the assessment still reduces to a single interpretive choice about which data cut to believe, and it leaves the confidence that choice warrants ungraded. The principal gap in the report is the same as in every companion case: the absence of an explicit, reproducible link from documented uncertainty to a stated degree of certainty, which is made acute here by a genuinely borderline survival signal.

CLEVER ACCESS
Let's be clever together

ANALYTICAL REVIEW

Clever-Access HTA Perspectives

Is a series of strategic reviews and expert analyses exploring the evolving landscape of European Health Technology Assessment, evidence generation, market access and healthcare policy.

DISCUSS WITH CLEVER-ACCESS ABOUT JCA AND/OR JSC, CONTACT:

Aleksandra Caban
aleksandra.caban@clever-access.com

FOR ANY QUESTIONS ABOUT THIS REPORT, CONTACT:

Lylia Chachoua
lylia.chachoua@clever-access.com

Interested in EU HTA, Market Access and Evidence Strategy?

Join the **Market Access Society** community for expert discussions, training programmes, webinars and practical resources designed for professionals working across the healthcare ecosystem.

SCAN CODE QR TO JOIN:

MAS COMMUNITY

FOR ANY QUESTIONS ABOUT THIS REPORT, CONTACT:

Mondher Toumi
mto@inovintell.com

VISIT OUR WEBSITE:

www.marketaccesssociety.org

Published by
MAS