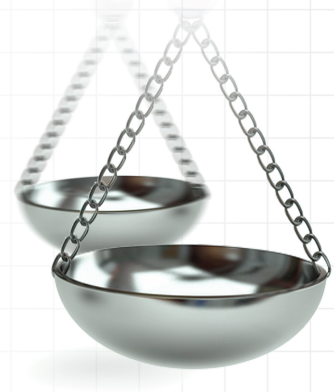
Industry Update

For reprint orders, please contact: info@becaris.com



Journal of **Comparative Effectiveness Research**

R WE ready for reimbursement? A round-up of developments in real-world evidence relating to health technology assessment: part 28

Paul Arora[1,2], Kirk Geale[3] & Sreeram V Ramagopalan*[,4]

[1]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, M5T 3M7, Canada
[2]Inka Health, 100 King St W. Suite 1600. Toronto, ON, M5X 1G5, Canada
[3]Athagoras, Gerbergasse 4, Basel, CH-4001, Switzerland
[4]Quantify Research, Hantverkargatan 8, Stockholm, 112 21, Sweden
[5]Centre for Pharmaceutical Medicine Research, King's College London, London, SE1 9NH, UK
*Author for correspondence: sreeram@ramagopalan.net

In this update, we consider the role of real-world evidence in the European Union's new Joint Clinical Assessment (JCA) process and also review a systematic review and meta-analysis of the concordance between target trial emulations and their benchmarking randomized controlled trials.

First draft submitted: 29 July 2026; Accepted for publication: 31 July 2026; Published online: 20 August 2026

Keywords: EU Health Technology Assessment Regulation • European Medicines Agency • health technology assessment • Joint Clinical Assessment • PICO • randomized controlled trial • real-world data • real-world evidence • target trial emulation

The European Union's Joint Clinical Assessment (JCA), established under Regulation (EU) 2021/2282, became operational in January 2025 and represents a significant change to the European health technology assessment (HTA) landscape [1–4]. For eligible centrally authorized medicines – beginning with oncology products and advanced therapy medicinal products – a single joint assessment of relative clinical effectiveness and safety is now conducted at EU level, running in parallel with the EMA marketing authorization review. The JCA covers only the clinical assessment; economic evaluation, pricing and reimbursement decisions remain with national HTA bodies, and the JCA report informs but does not replace those national appraisals. The mechanics matter for evidence planning: after EMA notification of an upcoming application, an assessor and co-assessor are appointed and a harmonized list of Population, Intervention, Comparator, Outcomes (PICO) questions is agreed across Member States. Only once that PICO list is finalized does the health technology developer (HTD) have 100 days to compile and submit the JCA dossier. Because comparators are drawn from national standards of care that differ across Member States, a single product could attract a large and heterogeneous set of PICOs, and the HTD does not know their content in advance.

Taylor and Adamson have set out where RWE could contribute under these conditions [5]. Their proposals fall into three broad categories. First, anticipating the PICOs: they advocate a structured 'PICO-forecast matrix' that identifies plausible comparators by Member States from current guidelines and standards of care, scores candidate PICOs by likelihood and market importance, and assesses for each high-priority PICO whether pre-existing real-world data (RWD) can supply the required population, comparator and outcomes within the 100-day window, flagging those that would need *de novo* evidence. Second, addressing PICOs directly, most obviously through external comparator arms constructed from RWD where a single-arm trial leaves a comparator undefined. Third, supporting the analysis: RWD platforms can be used ahead of time to identify and characterize prognostic factors and effect modifiers, so that transportability analyses or population-adjusted indirect comparisons can be prespecified rather than constructed *post hoc*, and doubly robust methods (for example propensity score weighting combined with outcome regression) can be applied to guard against violations of modelling assumptions. Underlying everything

is that under a 100-day clock, evidence generation must move from reactive to proactive, because the work cannot be started once the PICOs are known.

It is now possible to look at what has actually happened. We reviewed the three published JCA reports for medicinal products published at the time of writing – tovorafenib (pediatric low-grade glioma), lurbinectedin (extensive-stage small cell lung cancer [ES-SCLC]) and tarlatamab (ES-SCLC) [6–8] – examining where RWE was used, where evidence was not submitted at all, and where the indirect comparisons that were submitted attracted criticism from the assessors.

The first observation is that RWE is almost entirely absent. Across the three reports, 16 PICOs were defined, and not one accepted a comparative effectiveness estimate that was derived from RWE. The single genuine attempt to use RWD occurred in the tovorafenib assessment. The scope defined eight PICOs across three populations, and the pivotal evidence came from a single-arm study (FIREFLY-1), so a comparator had to be sourced externally for every PICO. Several PICOs specified an 'individualized treatment' comparator comprising multiple therapies. The HTD conducted a feasibility study to identify RWD sources from which external control arms could be constructed; one US database initially appeared feasible but was ultimately deemed unsuitable owing to an insufficient number of patients with BRAF alterations and inadequate follow-up in the relapsed/refractory setting, and the HTD concluded that a fit-for-purpose comparator data source could not be found.

The second observation concerns the scale of the resulting evidence gaps. For tovorafenib, six of the eight PICOs were left with no comparative data whatsoever. The assessors faulted the HTD's search strategy as unduly restrictive: the HTD sought comparator studies only where the comparator included every treatment option comprising the individualized strategy, whereas data covering a relevant selection of those options might have been assessable had they been submitted. One PICO fell away on documentation grounds – the unanchored matching-adjusted indirect comparison (MAIC) for PICO 7 (versus trametinib) was excluded because the comparator study, TRAM-01, was reported only in conference abstracts, leaving insufficient information to assess its methods, results or patient characteristics. In the end, a single PICO out of eight (PICO 5, vs dabrafenib plus trametinib) reached assessment. Tarlatamab and lurbinectedin fared better on completeness: the lurbinectedin scope defined a single PICO, addressed directly by the head-to-head Phase III IMforte trial in which the comparator (atezolizumab monotherapy) sat within the randomization. The tarlatamab scope defined four populations and seven PICOs, and the HTD submitted evidence for all seven, drawing on the pivotal randomized trial DeLLphi-304.

The third observation is that where non-randomized comparisons were submitted, the assessors criticized them heavily – and that these comparisons were built on published trial data rather than RWD. For tovorafenib's PICO 5, the unanchored MAIC re-weighted FIREFLY-1 against the published Bouffet 2023 study. The assessors judged that three of the four key MAIC assumptions may not hold, and that the fourth (the covariance structure of baseline variables in the comparator) could not even be evaluated because the comparator publication did not report it, making this an additional source of bias by default. The small effective sample size limited the number of prognostic and effect-modifying variables that could be adjusted for, producing imprecision, and the assessors raised concerns about the comparability of outcome definitions. For tarlatamab, the picture is similar but more elaborate. Direct comparison with topotecan was available for three PICOs; a Bayesian network meta-analysis (NMA) was used for two; an unanchored MAIC against cyclophosphamide/doxorubicin/vincristine (CAV) for one; and for the last, only a single-outcome indirect comparison. The comparator inputs, however, were published trials – GFPC 0501 for the MAIC, and Baize 2020 and von Pawel 1999 for the NMA. The assessors concluded that the similarity assumption between DeLLphi-304 and Baize 2020 was 'severely violated' and rated the certainty of the NMA-based indirect comparisons as very low, citing confounding of overall survival by subsequent therapies, open-label designs in both studies, and differing definitions of overall survival between DeLLphi-304 and von Pawel 1999 (time from randomization vs time from first administration). The networks lacked closed loops, so inconsistency could not be assessed, and for von Pawel 1999 the confidence interval had to be reconstructed by extracting pseudo-individual patient data from a published Kaplan–Meier curve.

Could RWE have played a stronger role if it had been planned for? For lurbinectedin the answer is plainly no. Where the pivotal trial's comparator coincides with the assessed PICO, the heterogeneous-PICO problem does not arise and RWE has little to add to the relative effectiveness assessment. For tovorafenib the answer is a qualified yes. The proximate cause of six unaddressed PICOs was the absence of a usable comparator dataset, and the feasibility study appears to have been conducted reactively, in response to the PICOs, at which point the only available answer was that no adequate source existed. The commentary's proposals map directly onto this failure: an RWD landscape assessment, conducted before EMA notification, would have established well in advance whether any data source

could supply BRAF-altered relapsed/refractory pediatric LGG patients with adequate follow-up, and if so, would have allowed data access, governance and analytical templates to be put in place ahead of the clock starting. The qualification is that a proactive strategy might well have concluded that no adequate RWD source exists for this population anywhere – this is a rare pediatric cancer, and the HTD's conclusion may simply be correct. But that conclusion would then have been reached early enough to act on, whether by seeking linked or registry data, by prospective data collection, or by adjusting the development plan – rather than being documented as a dead end in the dossier.

Tarlatamab is the most interesting case, because here the evidence gaps are not gaps in submission but weaknesses in what was submitted, and they are weaknesses of a kind RWE is well suited to address. The assessors' objections to the NMA were not primarily statistical: they concerned the comparator evidence itself. A trial published in 1999 was used to inform the effectiveness of a comparator in a 2025 assessment, with an outcome definition that did not match the pivotal trial, a Kaplan–Meier curve that had to be digitized because the underlying estimates were unavailable, and an inability to account for the subsequent therapies that confound overall survival in a contemporary treatment pathway. A well-constructed real-world cohort of ES-SCLC patients receiving CAV or platinum re-challenge in current European practice would have addressed several of these problems at once: individual patient data would be available, outcome definitions could be aligned to those of DeLLphi-304, subsequent therapies could be measured and adjusted for, and the comparator population would reflect the standard of care that Member States are actually asking about rather than the practice of a quarter of a century ago. This is precisely the use case the commentary describes – using RWD to inform or address PICOs where the published comparator evidence is old, thin, or misaligned – and it is the one place across these three reports where a planned RWE strategy might have materially improved the certainty of the assessment.

A necessary caveat to all of the above is that RWE would only strengthen a JCA dossier if it were good enough to be believed. Target trial emulation (TTE) – in which an observational analysis is explicitly designed to mimic the protocol of a hypothetical or actual randomized controlled trial (RCT), specifying eligibility, treatment strategies, assignment, time zero, outcomes and the estimand up front in order to avoid design-related biases such as immortal time bias – has become the dominant framework for generating credible comparative effectiveness evidence from RWD, and is increasingly endorsed in HTA and regulatory methods guidance [9,10]. Wang and colleagues have published the most comprehensive empirical test of the approach to date: a systematic review and meta-analysis of the concordance between explicitly reported target trial emulations and their benchmarking RCTs [11]. Across 107 TTE–RCT pairs from 50 studies, the overall summary ratio of ratios was 0.96 (95% CI: 0.92 to 1.01) and 92 of 107 pairs (86%) fell within the pre-defined concordance range, indicating no systematic difference in effect estimates overall. However, the Pearson correlation between paired estimates was only 0.59 (95% CI: 0.45–0.70), consistent with moderate rather than strong agreement, with meaningful between-pair heterogeneity.

The findings that matter most concern what drove concordance. In the 63 pairs judged to be 'close emulations' of the trial design, agreement rose substantially (Pearson correlation: 0.83). Emulations built on linked data, registries and multicountry databases agreed with their trials more closely than those built on standalone claims or electronic health record data, and claims-based emulations tended to underestimate treatment effects (ratio of ratios: 0.90; 95% CI: 0.82–0.99). Discordance was systematically greater when treatment was initiated in hospital (poorly captured in claims), when baseline populations were imbalanced in age or sex, and when outcome emulation quality was low – for example, administrative definitions of venous thromboembolism, or laboratory-based end points with high missingness. Given that TTE is endorsed by HTA agencies, the temptation is to read the headline null result as a general validation of observational comparative effectiveness. It is not. The design features associated with agreement (fidelity to the target trial protocol, careful population matching, specific and completely captured outcomes, rich linked or registry data) are the same features HTA assessors scrutinize, which means the levers that make an emulation agree with a trial are the levers that make it credible to a payer.

This sets a demanding but achievable bar for the RWE opportunities identified above. For tovorafenib, it supports the view that a thin, poorly matched external control arm would not have rescued the unaddressed PICOs – the assessors criticized even the trial-derived MAIC for unadjusted effect modifiers, limited overlap and imprecision, and an RWD-based comparison with the same weaknesses would have met the same fate. For tarlatamab, by contrast, the Wang findings are encouraging: a comparator cohort drawn from linked or registry data, with outcomes aligned to the pivotal trial and subsequent therapies captured, sits squarely in the category of emulation design that reproduces randomized results well. The lesson is not that RWE should be generated for every PICO, but that it

10.57264/cer-2026-0167

should be generated where the design conditions for credibility can be met – and that establishing whether they can be met is itself a task for the planning phase, not the submission phase.

The early JCA experience therefore offers a more sober picture than the enthusiasm surrounding RWE's role might suggest, but not a discouraging one, given the potential of high-quality RWE being able to generate rigorous results. RWE has so far contributed little to the first three JCA reports; the one attempt to deploy it failed on data availability before any analysis could be assessed; and the nonrandomized comparisons that were submitted, all built on published trial data, were received critically. Yet the reports also show exactly where a planned RWE strategy would have earned its place: in supplying a contemporary, individual-level comparator cohort where the published evidence is decades old and misaligned, and in establishing early – rather than at the eleventh hour – whether a viable comparator dataset exists at all. For manufacturers, this reinforces that the value of RWE in HTA is realized through early, prospective planning and rigorous execution on fit-for-purpose data. Under the JCA, where the required evidence is not knowable until the clock has already started, that principle stops being good practice and becomes a precondition.

Financial disclosure

Author SV Ramagopalan has received an honorarium from Becaris Publishing for the contribution of this work. The authors have received no other financial and/or material support for this research or the creation of this work apart from that disclosed.

Competing interests disclosure

The authors have no competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Writing disclosure

No writing assistance was utilized in the production of this manuscript.

References

1. Ramagopalan SV, Pannelay AJ. Access in all areas? A round-up of developments in market access and health technology assessment: part 14. *J. Comp. Eff. Res.* 15(5), e260056 (2026).
2. Gilardino R, Treharne C, Mardiguian S, Ramagopalan SV. Access in all areas? A round up of developments in market access and health technology assessment: part 1. *J. Comp. Eff. Res.* 12(10), e230129 (2023).
3. Ramagopalan SV, Pannelay AJ. Access in all areas? A round-up of developments in market access and health technology assessment: first Joint Clinical Assessment report released. *J. Comp. Eff. Res.* 15(7), e260117 (2026).
4. Pannelay AJ, Gilardino RE, Ramagopalan SV. Access in all areas? A roundup of developments in market access and health technology assessment: part 6. *J. Comp. Eff. Res.* 14(3), e240239 (2025).
5. Taylor L, Adamson B. RWE Submission for European regulators and payers: challenges, uncertainties, and opportunities. *Ther. Innov. Regul. Sci.* 60(4), 929–933 (2026).
6. European Commission. Joint Clinical Assessment report on tarlatamab (Imdylltra). https://health.ec.europa.eu/publications/joint-clinical-assessment-report-tarlatamab-imdylltra_en
7. European Commission. Joint Clinical Assessment report on lurbinectedin (Zepzelca). https://health.ec.europa.eu/publications/joint-clinical-assessment-report-lurbinectedin-zepzelca_en
8. European Commission. Joint Clinical Assessment report on tovorafenib (Ojemda). https://health.ec.europa.eu/publications/joint-clinical-assessment-report-tovorafenib-ojemda_en
9. Arora P, Ramagopalan SV. R WE ready for reimbursement? A round-up of developments in real-world evidence relating to health technology assessment: part 21. *J. Comp. Eff. Res.* 14(11), e250148 (2025).
10. Bray BD, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 14. *J. Comp. Eff. Res.* 13(1), e230189 (2024).
11. Wang C, Tang D, von Dadelszen P *et al.* Concordance between target trial emulation and randomised controlled trials: systematic review and meta-analysis. *BMJ* 393, e086810 (2026).