



R WE ready for reimbursement? A round-up of developments in real-world evidence relating to health technology assessment: part 23

Journal of **Comparative Effectiveness Research**

Paul Arora^{1,2} & Sreeram V Ramagopalan^{*,3}

¹Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5T 3M7, Canada

²Inka Health, 100 King St W. Suite 1600. Toronto, ON M5X 1G5, Canada

³Centre for Pharmaceutical Medicine Research, King's College London, London, SE1 9NH, UK

*Author for correspondence: sreeram@ramagopalan.net

In this update, we explore a review on transportability methods to enable the use of cross-jurisdictional evidence when local data are limited, a review of clinical trials that use pragmatic elements and finally, we discuss a study highlighting the potential transformative role of large language models in disease progression modeling.

First draft submitted: 23 October 2025; Accepted for publication: 24 October 2025; Published online: 15 December 2025

Keywords: artificial intelligence • clinical trials • health technology assessment • large language models • pragmatic trial design • real-world data • real-world evidence • transportability

A fundamental challenge in health technology assessment (HTA) is the limited availability of local real-world data (RWD), particularly for rare diseases. Manufacturers may possess robust datasets from other jurisdictions but face questions about their relevance to local decision-making contexts. Gupta *et al.* have written a primer on methodology to address this challenge – transportability – and its implications for HTA [1]. The authors systematically outline the methodological framework for transportability, distinguishing it from the related concept of generalizability. While generalizability concerns whether study findings apply to the broader population from which participants were selected, transportability involves applying treatment effect estimates to a target population that may be partially or entirely distinct from the study population. This distinction becomes particularly relevant when manufacturers seek to leverage RWD from one jurisdiction to inform HTA submissions in another.

Gupta and colleagues note three foundational assumptions necessary for valid transportability: consistency (patients in different settings receive equivalent versions of treatment), positivity (all relevant patient subgroups in the target population are represented in the source data), and conditional exchangeability (all factors that differ between populations can be measured and adjusted for). The authors illustrate how violations of these assumptions manifest in real-world scenarios. For instance, consistency may be violated when subsequent therapy availability differs across jurisdictions, as different treatment landscapes can affect observed outcomes even if the index treatment is administered identically. Positivity fails when certain patient subgroups exist only in the target population but not in the source data; for example, if racial or ethnic groups present in the local context are absent from nonlocal studies. Conditional exchangeability is threatened when important prognostic factors like socioeconomic status remain unmeasured or are captured inconsistently across settings. The primer describes several statistical methods to address these assumptions, including matching, weighting, standardization and doubly robust methods. When important variables remain unmeasured, quantitative bias analysis can help by exploring how unmeasured or mismeasured variables could impact study findings through tipping point analyses or simulation-based approaches [2,3].

To illustrate how transportability concerns manifest in practice, Gupta *et al.* conducted a targeted review of Canada's Drug Agency (CDA-AMC) and NICE appraisals between 2015 and 2024 that included criticisms related to the use of nonlocal real-world evidence (RWE). Teclistamab for relapsed/refractory multiple myeloma exemplifies how a number of transportability assumptions can be simultaneously violated. The submission utilized

a US external control arm to support a single-arm trial, but CDA-AMC reviewers raised multiple concerns. Regarding consistency, the therapies used in the US control cohort were not considered relevant to Canadian standard of care, and the comparator data were deemed outdated (2018–2021 data for a 2024 assessment). The exchangeability assumption was questioned due to concerns about unmeasured confounding factors that might differ between US and Canadian settings. This case exemplifies how transportability issues compound when data are both geographically distant and temporally misaligned with the target decision context. Critically, the reviewed HTA submissions did not proactively apply formal transportability methods to address these concerns. The manufacturers submitted nonlocal RWE without systematically evaluating or adjusting for foundational assumptions, leading to rejection or substantial criticism from reviewers. This represents a missed opportunity. Had transportability methods been applied prospectively, manufacturers could have adjusted US control arm data to reflect Canadian population characteristics and treatment patterns and employed quantitative bias analysis to bound the potential impact of unmeasured confounders.

For pharmaceutical manufacturers, this gap between available methodology and current practice represents both a challenge and an opportunity. First, transportability should be considered during evidence planning rather than as a post hoc analytical challenge. When nonlocal data are anticipated to be necessary, study designs should prioritize collecting information on variables likely to differ across jurisdictions and that might confound treatment effects – including subsequent therapy patterns, adherence measures, socioeconomic indicators and quality-of-life assessments. Second, manufacturers should proactively apply the transportability methods described by Gupta *et al.* rather than simply presenting unadjusted nonlocal data and hoping reviewers will accept its relevance. Third, transparency about assumption violations and their potential impact is essential. Fourth, when possible, combining data from multiple jurisdictions rather than relying solely on a single nonlocal source may enhance credibility by demonstrating consistent patterns across contexts. The authors note that while transportability methods are advancing rapidly, their adoption in HTA remains limited. However, the consistent rejection of nonlocal evidence that lacks formal transportability analysis suggests that the status quo – submitting unadjusted nonlocal data – is increasingly untenable. Manufacturers who invest in applying transportability methods prospectively, rather than waiting for reviewers to identify these limitations, will be better positioned for successful HTA outcomes.

While transportability methods address analytical challenges with existing data, pragmatic trial designs offer an approach to generate evidence that more closely reflects real-world clinical practice. Traditional randomized controlled trials (RCTs), while considered the gold standard for establishing efficacy, often face limitations in reflecting real-world clinical practice. Highly selective patient populations, controlled settings, and limited long-term follow-up can constrain the generalizability of findings to diverse patient populations encountered in routine care. Clinical trials with pragmatic elements represent a middle ground between traditional explanatory RCTs and purely observational studies, incorporating design features that more closely resemble routine clinical practice while maintaining the rigor of randomization. The Pragmatic-Explanatory Continuum Indicator Summary-2 (PRECIS-2) framework has provided researchers with a systematic approach to evaluating and integrating pragmatic elements across nine domains, including eligibility criteria, recruitment, setting, treatment flexibility, follow-up and outcomes. Health authorities globally, including the US FDA and the EMA, have recognized the potential value of pragmatic trials and RWD in complementing traditional clinical evidence. Despite this growing regulatory interest, limited literature has systematically characterized the design features and RWD utilization patterns in pragmatic trials, representing a significant knowledge gap that Su and colleagues sought to address [4]. Su *et al.* conducted a targeted review focused on identifying clinical trials with pragmatic elements during trial design and conduct, as well as clinical trial extension studies utilizing RWD for long-term follow-up beyond the initial parent trial phase.

The review identified 27 use cases, comprising 22 clinical trials with pragmatic elements and five extension studies utilizing RWD. Regarding pragmatic elements in clinical trials, the most commonly implemented features included broad eligibility criteria (77.3% of studies), flexible treatment management allowing physicians to adjust treatments according to usual care practices (63.6%), minimal or no protocol-mandated follow-up visits (40.9%), and streamlined end point collection (45.5%). Only four trials systematically evaluated their pragmatic features using the PRECIS-2 tool, with total scores ranging from 34.3 to 41 out of a maximum of 45, indicating substantial pragmatism across multiple domains. Half of the identified pragmatic trials incorporated RWD, utilizing diverse data sources including electronic health records (EHRs), claims databases, and registries. The authors identified two primary RWD utilization patterns. First, four studies enriched trial data by integrating primary trial data with secondary databases to create enhanced clinical trial databases with more comprehensive patient information. Second, seven studies embedded trials within existing routine healthcare databases and systems, leveraging RWD

infrastructure for multiple trial processes including participant recruitment, screening, randomization, baseline data collection, outcome ascertainment and follow-up. The ADAPTABLE trial exemplified sophisticated RWD integration, utilizing EHRs to identify and recruit over 15,000 patients across 40 clinical sites for a pragmatic comparison of two aspirin doses in atherosclerotic cardiovascular disease. The trial employed EHR prompts, prescreening and electronic informed consent to facilitate efficient enrollment during outpatient visits, while incorporating both EHRs and Medicare claims data to capture end points including death, hospitalizations, coronary revascularization procedures and patient-reported outcomes. Event ascertainment demonstrated high reliability, with 92–100% of events identified through either EHRs or claims data, while the trial's cost was approximately a fifth to a half that of a traditional RCT of similar scale.

For the five extension studies utilizing RWD, three were completed studies with published results and two were ongoing studies identified through ClinicalTrials.gov. Two studies examined human papillomavirus (HPV) vaccine effectiveness using Nordic national registries linked via personal identification numbers, tracking incidence rates of HPV-related conditions including cervical intraepithelial neoplasia, adenocarcinoma *in situ*, and cervical, vulvar and vaginal cancers over extended periods up to 14 years. A respiratory syncytial virus extension study in the US employed tokenization, a privacy-preserving technique where patient identifiers are converted into unique, irreversible codes through algorithms, enabling researchers to link clinical trial participants to real-world healthcare databases without directly sharing personal information between systems [5]. This approach enables efficient tracking of long-term outcomes including hospitalizations and mortality through passive data linkage, particularly valuable in the fragmented US healthcare system that lacks standardized national patient identifiers. Notably, one extension study demonstrated a novel application of RWD for constructing external comparators rather than solely for long-term follow-up. The study evaluated three formulations of paliperidone palmitate for schizophrenia relapse prevention, drawing the intervention cohort from a single-arm extension study that enrolled patients rolling over from a Phase III trial, while deriving comparator cohorts from the IBM MarketScan Multistate Medicaid Database using propensity score matching.

For pharmaceutical manufacturers, the study findings present several strategic considerations for drug development and evidence generation programs. Su and colleagues note several ongoing regulatory initiatives supporting pragmatic trial development. The US FDA's Project Pragmatica aims to enable more pragmatic, patient-centric trials in oncology, while the FDA's Center for Drug Evaluation and Research has launched demonstration projects to partner with sponsors on planning pragmatic trials. Pragmatic trial designs offer the potential to generate clinically relevant evidence in populations and settings that more closely reflect real-world practice, potentially strengthening the generalizability and applicability of findings for regulatory and HTA submissions. The integration of RWD infrastructure into trial design can potentially substantially streamline recruitment, reduce protocol burden and facilitate efficient long-term follow-up – addressing common challenges in traditional trial execution. Further, extension studies utilizing RWD offer manufacturers opportunities to generate long-term safety and effectiveness data that address HTA bodies' requirements. However, pragmatism in trials may introduce challenges in internal validity through factors such as open-label designs, flexible treatment protocols and variable adherence patterns. HTA bodies must develop frameworks for assessing the acceptability of design trade-offs, recognizing that the optimal balance between internal and external validity depends on the specific decision context, available alternative evidence and clinical question being addressed. Collaborative efforts between manufacturers, regulators, HTA bodies and researchers to develop methodological standards and reporting guidelines will be essential to ensure that pragmatic trials and RWD integration fulfil their promise of generating robust, relevant evidence that supports informed decision-making and ultimately improves patient access to beneficial therapies.

Beyond innovations in trial design and data transportability, the emergence of large language models potentially represents a paradigm shift in how we might generate and synthesize RWE for HTA. We have previously discussed their application to the development of health economic models [6]. Delphi-2M demonstrates how generative pretrained transformer architectures – similar to those underlying ChatGPT – can be adapted to model disease progression [7]. Trained on health records from 0.4 million UK Biobank participants, Delphi-2M predicts rates for more than 1000 diseases conditional on individual health histories, achieving accuracy comparable to existing single-disease models. Critically, the model's generative capabilities enable sampling of synthetic future health trajectories up to 20 years ahead, providing meaningful estimates of disease burden and enabling training of AI models without exposure to actual patient data. External validation on 1.9 million Danish individuals demonstrated the model's generalizability across healthcare systems. For manufacturers, such approaches could revolutionize evidence generation – from identifying patient cohorts and predicting treatment responses to simulating long-term outcomes

in rare diseases where traditional data collection is challenging. HTA bodies like NICE have begun establishing frameworks for artificial intelligence use, emphasizing the need for clinical validation, bias assessment and clear documentation of model limitations [8]. The potential of large language models to accelerate evidence generation must be balanced against requirements for reproducibility, explainability and clinical meaningfulness that underpin HTA submissions.

In conclusion, evidence generation strategies should consider the full range of methodological approaches now available [9–11], selecting tools appropriate to specific decision contexts rather than defaulting to traditional approaches. Importantly, these methodological advances do not eliminate fundamental requirements for rigorous causal inference – they expand the settings where such inference may be possible and the efficiency with which evidence can be generated. As HTA bodies develop more explicit guidance on acceptable applications of these methods, manufacturers who invest in appropriate methodological capabilities will be better positioned to generate compelling evidence supporting patient access to innovative therapies.

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.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>

References

1. Gupta A, Duffield S, Shephard C *et al.* Transportability of nonlocal real-world evidence and its relevance to health technology assessment: a primer. *J. Comp. Eff. Res.* 14(10), e250041 (2025).
2. Gupta A, Hsu G, Kent S *et al.* Quantitative bias analysis for single-arm trials with external control arms. *JAMA Netw. Open* 8(3), e252152 (2025).
3. Leahy TP, Kent S, Sammon C *et al.* Unmeasured confounding in nonrandomized studies: quantitative bias analysis in health technology assessment. *J. Comp. Eff. Res.* 11(12), 851–859 (2022).
4. Su L, Chen L, Betigeri S *et al.* Clinical trials with pragmatic elements: a review of use cases and real-world data utilization. *Clin. Pharmacol. Ther.* 118(6), 1350–1365 (2025).
5. Arora P, Ramagopalan SV. RWE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 20. *J. Comp. Eff. Res.* 14(9), e250113 (2025).
6. Castanon A, Tsvetanova A, Ramagopalan SV. RWE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 16. *J. Comp. Eff. Res.* 13(8), e240095 (2024).
7. Shmatko A, Jung AW, Gaurav K *et al.* Learning the natural history of human disease with generative transformers. *Nature* 647(8088), 248–256 (2025).
8. Arora P, Ramagopalan SV. RWE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 17. *J. Comp. Eff. Res.* 14(1), e240212 (2025).
9. Arora P, Ramagopalan SV. RWE 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. Arora P, Ramagopalan SV. RWE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 19. *J. Comp. Eff. Res.* 14(7), e250063 (2025).
11. Arora P, Ramagopalan SV. RWE ready for reimbursement? A round-up of developments in real-world evidence relating to health technology assessment: part 22. *J. Comp. Eff. Res.* 14(12), e250149 (2025).