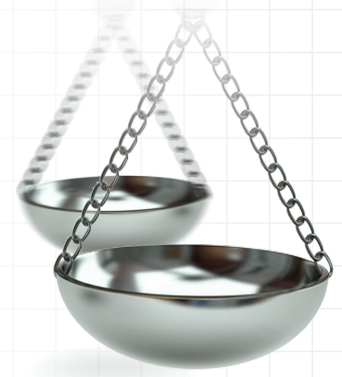




R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 18

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, Schwartz Reisman Innovation Campus, University of Toronto, Toronto, ON, M5G 0C6, Canada

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

*Author for correspondence: sreeram@ramagopalan.net

In this update, we discuss recent publications examining the use of real-world data as a measure of treatment effectiveness in submissions to the National Institute of Health and Care Excellence, the results of a pilot study from the Coalition to Advance Real-World Evidence initiative and a validation study of synthetic data.

First draft submitted: 5 February 2025; Accepted for publication: 7 February 2025; Published online: 18 February 2025

Keywords: CARE initiative • Coalition to Advance Real-World Evidence initiative • data privacy • external control arm • health technology assessment • NICE • oncology • randomized controlled trial • real-world data • real-world evidence • synthetic data • target trial emulation

The use of real-world data (RWD) for estimating treatment effects is one of the most exciting use cases of RWD as it could enable patient access to medicines where trials are infeasible; however, studies attempting to do this robustly require the highest methodological rigor (i.e., causal inference techniques) to avoid bias [1]. A recent analysis has attempted to investigate how RWD has been used to derive treatment effects for use in health technology assessment (HTA) by reviewing submissions to the National Institute for Health and Care Excellence (NICE) between January 2016 to December 2023 [2]. The review identified 64 HTA submissions (~11% of total submissions during this period) that used RWD to estimate treatment effects. The number of submissions employing RWD to measure treatment effects has increased over time, from 5% of all submissions in 2016 to 14% of all submissions in 2023. The main RWD sources used were disease registries and electronic health records (EHRs), with chart reviews being employed to a lesser extent. Cancer (hematologic and solid tumor) was by far the most common indication that RWD was used for (more than 80% of submissions). Most studies were international and multicenter with data primarily coming from the UK, USA, France and Germany. RWD was used in two main ways: to construct external control arms (ECAs) for comparison with single-arm trials or with randomized controlled trials (RCTs) where the control arm was inadequate (for example the control arm was not reflective of standard of care in the UK), and to inform long-term treatment effects through extrapolation of survival data. All 64 submissions reviewed used a RWD ECA. Non-UK RWD sources were used in 56% of submissions, primarily justified by small sample sizes in UK data or better survival end point availability. Non-UK RWD was not treated differently to UK RWD by committees. However, the suitability and quality of RWD sources were not formally assessed using tools like NICE's Data Suitability Assessment Tool [3]. Surprisingly, but similar to previously noted findings for submissions to the EMA [4,5], approximately a third of submissions still relied on unadjusted (naïve) comparisons. When adjustments were made, inverse probability weighting was the most common approach. In total 12 submissions used RWD (all from disease registries) to inform long-term treatment effects, mainly to calibrate parametric curve choice. The review identifies four key areas needing improvement to ensure that RWD is rigorously analyzed to meet the high bar of being able to be robustly used to measure treatment effects. The first concern is that many submissions still use naïve comparisons rather than population-adjusted treatment comparisons, this is something manufacturers

need to note and HTA agencies should further encourage the use of adjustment methods. Second, unmeasured confounding remains a major concern that receives little attention, and manufacturers should explore the potential impact of this, for example through quantitative bias analysis (QBA) [6]. The third issue is that the reporting of RWD studies lack detailed descriptions of study design and analytical assumptions; this can be addressed by adopting the target trial emulation framework, using a protocol template and requiring pre-registration of the study protocol, as has been described elsewhere [7,8]. Finally, there is poor alignment between single-arm trial designs and RWD-based ECAs, which can be improved by taking early steps in trial design to facilitate comparison with RWD, such as using more pragmatic inclusion criteria. This article aligns with articles discussed in previous issues of this series [4,9], with the ultimate take home message being that RWD submissions to regulators and HTA agencies are currently falling short of best practice being called out in their guidance. For manufacturers, these recent studies provide important insights when preparing for regulatory and/or HTA submissions, highlighting the need to follow guidance such as NICE's real-world evidence (RWE) framework [10] to apply more rigorous approaches to RWD study design and analysis. This may require additional investment in capabilities to ensure high-quality RWE generation but will ultimately be needed to ensure that RWD submitted meets the increasingly rigorous HTA standards.

Building on these insights about RWE quality, recent work from the Coalition to Advance Real-World Evidence (CARE) initiative provides valuable lessons about replicating trial results using RWD. The CARE initiative aims to better understand when RWD studies can provide valid conclusions about cancer treatment effectiveness, including identifying critical study design and data source characteristics, by systematically emulating RCTs in oncology. CARE recently published a pilot study to evaluate whether EHR data could emulate findings from the KEYNOTE-189 RCT examining pembrolizumab plus chemotherapy versus chemotherapy alone in metastatic non-small-cell lung cancer [11]. Using the TriNetX Dataworks EHR database, the study identified 1854 eligible patients in the US – 589 receiving pembrolizumab plus chemotherapy and 1265 receiving chemotherapy alone. The hazard ratio for mortality was 0.95 (95% confidence interval [CI]: 0.78–1.16) in the intention-to-treat analysis, compared with 0.49 (95% CI: 0.38–0.64) in the RCT. The results therefore did not achieve regulatory agreement (where the emulation study estimate has a similar direction and statistical significance as the RCT estimate) or estimate agreement (when the emulation study point estimate lies within the bounds of the 95% CI of the RCT estimate). The lack of regulatory or estimate agreement does not necessarily indicate a failure of the RWD study. The authors suggest several factors that may have contributed to these discordant results. The study was unable to operationalize key RCT eligibility criteria like performance status, relied on diagnosis codes rather than clinical documentation for metastatic disease timing (which may have impacted line of therapy assessments) and there was missing data on key prognostic factors. Treatment crossover was substantial, with approximately 11% of comparator patients receiving pembrolizumab. The researchers could have tried to quantify the impact of unmeasured confounders, mismeasurement and missing data elements on their results using QBA [6]. Had QBA been performed for known sources of bias, a closer alignment to the target trial could have been possible. Interestingly, in the subgroup with *de novo* metastatic disease, where tumor registry data informed initial metastasis date, results more closely aligned with the RCT, suggesting data availability impacts emulation success. This pilot contrasts with the findings from the RCT-DUPLICATE initiative, where 75% of the comparisons between the RCT and RWD replication met regulatory agreement and 66% met estimate agreement [12,13]. The RCT-DUPLICATE authors concluded that three factors were important in determining whether RWD could reliably replicate a RCT – the ability to emulate key features of the trial in the available RWD, residual confounding and chance [13]. Reasons for inability to emulate a trial in RWD in RCT-DUPLICATE included poorer adherence of medications in the real world and missingness of outcomes [13]. Despite using EHR data with more clinical information as compared with claims data used by RCT-DUPLICATE, the CARE initiative shows that data quality and completeness is perhaps more important for oncology emulation studies than the indications studied by RCT-DUPLICATE. The CARE study emphasizes the critical importance of data source selection, with sources needing to have reliable and complete capture of disease staging and progression, treatment patterns and discontinuation, and key prognostic factors. Future CARE initiative studies will more critically evaluate the ability of data sources to operationalize key study design elements before proceeding with trial emulation. Learnings from this study and future studies to come from the CARE initiative should help guide decision makers as to which data sources and which key variables allow for trial emulation and robust treatment effect estimation in oncology.

While improving RWD quality remains crucial, addressing data privacy concerns through synthetic data generation offers another promising avenue for advancing RWE. Data privacy issues may impede access to RWD and sharing RWD across regions. Synthetic data, the health records of 'realistic' patients generated using sophisticated

algorithms based on the statistical properties of real populations, may be a potential solution to this challenge but needs rigorous validation to demonstrate its utility for regulatory and HTA decision-making. A recent study has attempted to validate synthetic data by comparing the similarity and stability of results between computationally derived synthetic EHR data versus original EHR data [14]. The study replicated a published retrospective cohort study on the real-world effectiveness of the Pfizer/BioNTech COVID-19 vaccine by Maccabi Healthcare Services in Israel using synthetic data generated from the same database, and the results from synthetic RWD were compared with the original RWD. Synthetic data were generated using the MDClone platform, which computationally derives synthetic data retaining statistical properties of the source data but without containing any actual patient information. Five synthetic dataset replicates were created to assess result stability. The results demonstrated strong validity of the synthetic data across multiple dimensions. Distribution of demographic and clinical characteristics showed very small differences between synthetic and original data, with standardized mean differences below 0.01. Vaccine effectiveness analyses demonstrated 100% regulatory and estimate agreement. For hazard ratios of COVID-19-related hospitalization and odds ratios for symptomatic COVID-19 infection, statistical tests showed no significant differences between synthetic and original estimates but regulatory and estimate agreement was not 100% for some subgroups. This study is one of only a few demonstrating the validity of synthetic data and the ability to preserve complex relationships in data necessary for effectiveness analyses is encouraging. The analytical approach may provide some best practices for synthetic data generation such as generating multiple replicates. The growing importance of RWE for decision-making combined with increasing data privacy challenges suggest synthetic data will likely play a larger role in future HTA submissions. While no HTA guideline to date has provided a perspective on synthetic data, in the US legal guidelines are beginning to emerge as to how synthetic data should be treated [15]. However, the bar for acceptance of synthetic data may be even greater than that for original RWD. Manufacturers will need to carefully validate synthetic data approaches and transparently document methods. Cross-sector collaboration will be needed to establish standards for synthetic data quality, validity assessment and privacy protection. While this validation study provides an important framework for evaluating synthetic data validity, additional demonstration projects across different therapeutic areas, data types and use cases are still needed. As synthetic data methods evolve, maintaining transparency and establishing clear standards will be critical for appropriate adoption in regulatory and HTA settings.

The evolving landscape of RWE in HTA submissions highlights both progress and challenges. While the increasing use of RWD for treatment effect estimation is encouraging, manufacturers must raise their analytical standards to meet HTA requirements. As novel data sources and methods emerge, continued collaboration between industry, HTA agencies and researchers will be required to advance their use in HTA.

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 financial and/or nonfinancial competing interests or relevant affiliations with any organization/entity to declare that are relevant to the subject matter or materials discussed in this manuscript. This includes employment, grants or research funding, consultancies, membership on scientific or other advisory boards, honoraria, stock ownership or options, paid expert testimony, 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. Ramagopalan SV, Simpson A, Sammon C. Can real-world data really replace randomised clinical trials? *BMC Med.* 18, 1–2 (2020).
2. Che Y, Duffield S, Gomes M. The use of real-world data for estimating relative treatment effects in NICE health technology assessment submissions: a review. *Pharmacoeconomics* 43(2), 123–131 (2025).

3. NICE. NICE real-world evidence framework. Appendix 1: Data Suitability Assessment Tool (DataSAT). (2022). Available at: <https://www.nice.org.uk/corporate/ecd9/chapter/appendix-1-data-suitability-assessment-tool-datasat>
4. Arora P, Ramagopalan SV. R WE 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).
5. Hogervorst MA, Soman KV, Gardarsdottir H, Goettsch WG, Bloem LT. Analytical methods for comparing uncontrolled trials with external controls from real-world data: a systematic literature review and comparison with European regulatory and health technology assessment practice. *Value Health* 28(1), 161–174 (2025).
6. 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).
7. 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).
8. Castanon A, Bray BD, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 15. *J. Comp. Eff. Res.* 13(5), e240033 (2024).
9. Castanon A, Duffield S, Ramagopalan S, Reynolds R. Why is target trial emulation not being used in health technology assessment real-world data submissions? *J. Comp. Eff. Res.* 13(8), e240091 (2024).
10. NICE. NICE real-world evidence framework. Overview (2022). Available at: <https://www.nice.org.uk/corporate/ecd9/chapter/overview>
11. Merola D, Campbell U, Lenis D *et al.* Calibrating observational health record data against a randomized trial. *JAMA Netw. Open* 7(9), e2436535 (2024).
12. Bray B, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 12. *J. Comp. Eff. Res.* 12(7), e230092 (2023).
13. Wang SV, Schneeweiss Set *al.*; RCT-DUPLICATE Initiative *et al.* Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. *J. Am. Med. Assoc.* 329, 1376–1385 (2023).
14. Wang E, Mott K, Zhang H, Gazit S, Chodick G, Burcu M. Validation assessment of privacy-preserving synthetic electronic health record data: comparison of original versus synthetic data on real-world COVID-19 vaccine effectiveness. *Pharmacoepidemiol. Drug Saf.* 33(10), e70019 (2024).
15. Utah AI. Laws require consumer-facing disclosures starting May 1. Available at: <https://www.orricks.com/en/Insights/2024/04/Utah-AI-Laws-Require-Consumer-Facing-Disclosures-Starting-May-1>