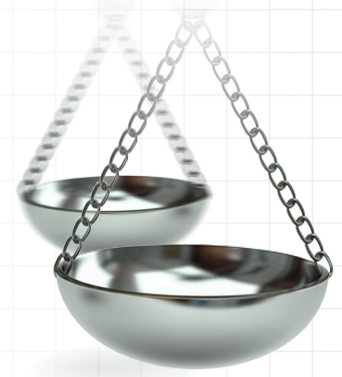




R WE ready for reimbursement? A round-up of developments in real-world evidence relating to health technology assessment: part 24

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 review a new bias appraisal tool, explore lessons from a trial emulation study, and describe the development of real-world evidence guidance in the Philippines.

First draft submitted: 14 January 2026; Accepted for publication: 23 January 2026; Published online: 9 February 2026

Keywords: bias • clinical trials • health technology assessment • methodology • Philippines • real-world data • real-world evidence • target trial emulation

Real-world evidence (RWE) has been employed for decades in health technology assessment (HTA) for evaluating disease epidemiology, healthcare resource utilization and costs. In recent years, HTA agencies have increasingly recognized the potential of RWE to inform decisions regarding treatment effects [1]. However, the utilization of RWE for evaluating treatment effects requires greater methodological scrutiny. The absence of randomization, incomplete capture of study variables and inappropriate study design or analytical approaches may produce biased findings. Practical tools to help HTA agencies evaluate whether RWE studies are sufficiently valid to influence decision making may therefore be helpful. Although HTA bodies have issued RWE guidance, such guidance frequently is at a high level and fails to provide systematic processes for comprehensive bias evaluation. A systematic review of assessment tools for non-randomized studies identified that no existing instrument covered all sources of bias known to affect RWE validity. Bykov and colleagues, working through the International Society for Pharmacoepidemiology (ISPE), addressed this need by developing Appraisal of Potential for Bias in Real-World Evidence Studies (APPRAISE), a comprehensive tool designed to guide bias assessment in observational studies evaluating medication comparative effectiveness or safety [2]. The multidisciplinary team working on APPRAISE comprised experts in pharmacoepidemiology, biostatistics, therapeutics and medicine, with HTA representatives included to ensure fitness-for-purpose for intended end-users. The development process incorporated multiple methodological approaches. A literature review was conducted to identify recently published tools, questionnaires, best practice documents and guidelines pertaining to RWE assessment. Additionally, a survey was distributed to HTA agencies across Europe, North America and the Asia-Pacific region to ascertain barriers to RWE utilization, currently employed assessment tools, and familiarity with sources of bias in observational studies of treatment effects. Responses were received from six agencies. Based on survey findings and internal deliberations, the development team elected to create a fillable Excel questionnaire utilizing binary response options. Notably, the developers deliberately avoided implementing a scoring system, recognizing that bias from different sources is context-specific, difficult to quantify and may not be cumulative. The tool evaluates potential bias across nine distinct domains. Study design biases encompass time-related bias (including immortal time), inappropriate adjustment for causal intermediaries, depletion of outcome-susceptible individuals (selection bias), reverse causation, detection bias and informative censoring. Misclassification bias addresses inaccurate measurement or missing data pertaining to exposures and outcomes. Confounding assessment considers design-based confounding control, adjustment for measured confounders and sensitivity analyses for residual confounding. Questions within each domain are

formulated such that affirmative responses indicate potential for bias. Response options include ‘unclear’ or ‘unsure’ for instances where reported data are insufficient or clinical expertise is required. Critically, responses auto-populate a summary evaluation of bias potential across all domains, accompanied by considerations for further action, including sensitivity analyses, study redesign recommendations or requests for additional statistical expertise. APPRAISE addresses several limitations identified in existing bias assessment tools and the authors explicitly position APPRAISE as complementary to existing instruments rather than a replacement. For example, they recommend concurrent utilization with HARPER for protocol harmonization.

The instrument provides HTA assessors with structured mechanisms to identify specific, avoidable sources of bias – particularly study design flaws that have historically proven difficult to detect. Ideally it would be used together with the FRAME recommendations to provide better reporting on the evaluation of RWE by HTA decision makers [3]. Manufacturers may also benefit from prospective utilization of APPRAISE during study planning to identify and prevent bias sources prior to study conduct. There is now a wealth of guidance documents on ‘what good RWE looks like’ [3–9], and they should all hopefully work to generally improve the quality of submitted RWE.

A recent example of guidance development comes from Tan-Lim and colleagues who describe the systematic development of the Philippine guidance for using RWE in clinical evaluation of health technologies [10]. The Philippines’ Universal HealthCare Act mandates HTA for all health technologies, yet the existing national HTA methods guide lacked specific provisions for RWE utilization. To address this gap, the research team conducted a comprehensive systematic review identifying 79 journal articles and nine international guidance papers covering RWE definitions, appropriate use cases, search and selection methods, critical appraisal approaches, data extraction procedures and synthesis techniques. This evidence base informed development of a draft guidance document subsequently validated through key informant interviews with international experts from Singapore, Thailand and Australia, followed by pilot testing with five potential users evaluating the guidance across diverse health technologies including oncology drugs, antibiotics and insulin products. A final review was conducted by the Philippine HTA Council and HTA Division. The expert consultation process yielded important refinements to the guidance. Experts emphasized that RWE should not automatically substitute for underpowered or poorly conducted RCTs, and may only be appropriate when RCTs are genuinely infeasible due to rare disease contexts or ethical constraints, or when RCTs cannot capture long-term end points. The experts also made additional suggestions around N-of-1 trials not constituting RWE, target trial emulation as an emerging methodological approach, qualitative research providing patient and caregiver experience evidence and the need to consider completeness and accuracy of secondary data. The pilot assessment confirmed the document’s logical organization and comprehensibility, while identifying areas requiring clarification. Key revisions included changing ‘noninterventional setting’ to ‘nonexperimental setting’ in the real-world data (RWD) definition to avoid misinterpretation, providing examples of surrogate outcomes and selection bias, and clarifying why adjusted effect estimates are preferred over raw data in analysis. The HTA Council and Division review resulted in additional refinements, including clarification that pragmatic clinical trials are considered RWE despite their experimental setting and addition of pandemics as emergency situations where RWE may guide decision-making pending RCT results.

The Philippine HTA environment now has explicit methodological standards for evaluating RWE submissions. The document’s emphasis on transparency, reproducibility and adherence to reporting guidelines draws substantially from frameworks from NICE, CDA-AMC and the FDA. The guidance will be updated in the future as needed. As noted above, the development of consistent guidance documents across agencies represents an important step toward establishing global consensus on what constitutes methodologically sound and acceptable RWE for HTA decision-making, providing manufacturers with clearer roadmaps for evidence generation.

While tools like APPRAISE and standardized guidance frameworks establish what methodologically sound RWE should look like, generating valid comparative effectiveness evidence from observational data can be challenging in practice, as illustrated by recent target trial emulation attempts. Target trial emulation – the approach of designing observational studies to mimic the structure of randomized controlled trials (RCTs) – has emerged as a methodological framework for addressing this challenge [3,11]. In this regard, Kong and colleagues undertook such an emulation study of the MONALEESA-2 trial, which evaluated ribociclib plus letrozole for advanced breast cancer, to both test the feasibility of trial emulation and assess whether relaxing certain entry criteria would alter treatment effectiveness estimates [12]. The researchers selected the MONALEESA-2 trial as an emulation target based on several factors favoring replication in RWD: the use of an active control arm (letrozole) rather than placebo, the outcome of overall survival (OS) which is objectively measurable in routine practice, and a sufficiently historical FDA approval date (March 2017) allowing adequate follow-up time. They conducted a comprehensive data feasibility

assessment using the SPIFD2 framework [11] to identify the most appropriate data source, ultimately selecting Optum's de-identified Market Clarity Data, which contains linked electronic health record (EHR) and claims data for over 1.8 million US patients. The study identified postmenopausal women with recurrent or metastatic breast cancer who initiated either ribociclib plus letrozole or letrozole alone between March 2017 (ribociclib approval date) and September 2023. The researchers systematically applied MONALEESA-2 inclusion and exclusion criteria where possible, excluding safety-related criteria. The original trial population comprised 668 participants (334 in each arm) and found a significant survival benefit with ribociclib (hazard ratio [HR] = 0.76, 95% CI: 0.63–0.93). To control for confounding in the RWD analysis, the researchers employed both inverse probability of treatment weighting (IPTW) and 1:1 propensity score matching, adjusting for demographic characteristics, performance status, biomarkers, comorbidity scores and prior treatments.

After applying trial eligibility criteria and baseline observability requirements, the sample was reduced to 3912 patients (106 ribociclib plus letrozole; 3806 letrozole alone) – representing only 3% of the initial population. Substantial differences emerged between trial participants and real-world patients and within the RWD population, patients treated with ribociclib plus letrozole had markedly higher comorbidity burdens and were much more likely to have metastatic disease compared with letrozole-only patients. These imbalances persisted even after propensity score adjustment, suggesting substantial unmeasured confounding. Critically, the researchers were unable to replicate the trial's survival benefit. Instead of the protective effect observed in MONALEESA-2, the crude RWD analysis yielded HR = 3.07, remaining above one even after adjustment (IPTW: HR = 2.75; propensity score matching: HR = 1.50). Multiple *post hoc* analyses, including use of historical control groups from before CDK4/6 inhibitor approvals and restriction to patients with documented metastatic disease, failed to produce results consistent with the trial findings. The failure of this emulation attempt stands in stark contrast to a successful emulation of MONALEESA-2 conducted using French registry data. The authors attribute this divergence primarily to data source differences. The French registry contained more complete clinical information, including ECOG performance status for nearly half of patients (versus ~10% in Optum), employed multiple imputation methods for missing data, and had physician-certified death reporting. Additionally, the registry provided longer follow-up (median 75 months) compared with the 27 months in the Optum data (15 months for ribociclib patients), allowing observation of the survival benefit that emerged after approximately 20 months in the trial.

This study yields several critical lessons for manufacturers considering target trial emulation. First, as noted by others [7], the authors emphasize that actual data exploration is necessary to understand the extent and patterns of missingness in key variables, which may differ substantially from general data source statistics. Variables crucial for oncology studies – including biomarker status, performance status and disease staging – showed high levels of missingness in this commercially available database, fundamentally limiting the ability to control for confounding. Further, outcome ascertainment completeness is paramount. The observed 7% absolute probability of death in this RWD analysis versus 54–66% in the trial indicates substantial under ascertainment, whether due to incomplete follow-up, deaths occurring outside hospital settings not captured in EHR/claims data, or data lag issues. Finally, the relatively recent approval of ribociclib (2017) combined with expected data lags in RWD sources meant that even with complete data through 2023, many patients lacked sufficient follow-up time to observe treatment effects. Second, the study demonstrates that trial emulation may not be feasible following approval of highly effective therapies that dramatically alter treatment landscapes. The introduction of CDK4/6 inhibitors between 2015 and 2017 created a 'game-changing' shift in advanced breast cancer treatment. Patients who did not receive these effective therapies post-approval likely differed systematically from those who did, making them invalid comparators for mimicking randomization. This failed emulation underscores that rigorous RWE methods require not only appropriate study design frameworks but also data sources with sufficient completeness to adequately control for confounding and ascertain outcomes.

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. Arora P, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 18. *J. Comp. Eff. Res.* 14(4), e250014 (2025).
2. Bykov K, Jaksa A, Lund JL, *et al.* APPRAISE: a tool for appraising potential for bias in real-world evidence studies on medication effectiveness or safety. *Value Health* 28(12), 1849–1856 (2025).
3. 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).
4. Bray BD, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 13. *J. Comp. Eff. Res.* 12(11), e230141 (2023).
5. Arora P, Ramagopalan SV. R WE 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).
6. 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).
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. Simpson A, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 8. *J. Comp. Eff. Res.* 11(13), 915–917 (2022).
9. Simpson A, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 4. *J. Comp. Eff. Res.* 11(1), 11–12 (2022).
10. Tan-Lim CSC, Cabaluna ITG, Infantado-Alejandro MAJ *et al.* Development of the Philippine guidance document for the use of real-world evidence for clinical assessment of health technologies. *Int. J. Technol. Assess. Health Care* 41(1), e72 (2025).
11. 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).
12. Kong AM, Andrean D, Khan S *et al.* Considerations for emulations of randomized controlled trials using real-world data: learnings from an emulation of MONALEESA-2. *J. Comp. Eff. Res.* 14(11), e250026 (2025).