Enhanced Randomised Controlled Trials-Real-World Evidence (RCT-RWE) Agreement Assessment Metrics for Health Technology Assessment (HTA)

MSR64
Poster presentation at
ISPOR Europe 2024
Nov 17-20, 2024, Barcelona

Jen-Yu Amy Chang[1,2], Saleema Rex[1], Jim Chilcott[1], Nicholas Latimer[1,3]

[1] Sheffield Centre for Health and Related Research (SCHARR), University of Sheffield, Sheffield, United Kingdom; [2]Lumanity, Sheffield, United Kingdom; [3]Delta Hat, Nottingham, United Kingdom

Objective

- To introduce enhanced metrics to evaluate how well real-world evidence (RWE) derived from Target Trial Emulations (TTEs) agrees with randomised controlled trials (RCTs) for benchmarking purposes, aiming to improve the assessment of RWE reliability for health technology assessment (HTA).
- While existing metrics focus on assessing the agreement of relative effects (e.g., risk ratio, RR) [1], the validity of absolute effects (e.g., survival) is equally crucial in HTA. We propose an enhanced metrics to address this.

Background

- ***Current landscape of RCT-RWE agreement assessment reporting***
  - A prior systematic review examined 97 studies using the TTE framework for estimating comparative effectiveness [2]. A subset of 14 studies focused on TTE benchmarking to closely emulate existing RCTs and thus were included to assess agreement methods for RCT and RWE concordance (Table 1).
  - Most of these 14 studies (11 studies, or 79%) used regulatory agreement as the basis for determining concordance.
  - Only two studies fully applied the standard three-criteria RCT-RWE agreement assessment matrix (i.e., purple-highlighted rows in Table 3) [1, 5], with one being the RCT DUPLICATE project that originally established it [1].
  - Three studies reported consistent estimates but gave no additional details [9, 10, 15], while one study compared survival curves [6].

Table 1. RCT-RWE Agreement Assessment in TTE Benchmarking Studies

Study reference	RCT-RWE Agreement Assessment			
	1. Regulatory agreement	2. Estimate agreement	3. Statistical agreement	Further documentation
Althunian et al. (2020) [3]	Yes	-	-	Non-inferiority conclusion
Petito et al. (2020) [4]	Yes	Yes	-	-
Yiu et al. (2020) [5]	Yes	Yes	Yes	-
Keyhani et al. (2020) [6]	Yes	-	-	Compared survival curves
Abrahami et al. (2021) [7]	Yes	Yes	-	-
Franklin et al. (2021) [1]	Yes	Yes	Yes	-
Matthews et al. (2021) [8]	Yes	-	-	-
Kirchgesner et al. (2021) [9]	-	-	-	Consistent estimates
Burn et al. (2019) [10]	-	-	-	Consistent estimates
Karaboyas et al. (2020) [11]	Yes	-	-	-
Admon et al. (2019) [12]	Yes	-	-	-
Lodi et al. (2019) [13]	Yes	-	-	-
Bacic et al. (2020) [14]	Yes	-	-	-
Boyne et al. (2021) [15]	-	-	-	Consistent estimates

- ***Caveats of the standard RCT-RWE Agreement Assessment Matrix for HTA***
  - Primarily assesses relative effectiveness agreement, but concordance of absolute effectiveness can be equally important for estimating state occupancy in economic evaluations, affecting cost and quality-of-life calculations.
  - The lack of explicit assessment of absolute effect agreement between RCT and RWE may overlook consistent immortal time bias across study arms (Figure 1a).
  - Lacks explicit survival curve concordance assessment, which can lead to incorrect conclusions, especially when assumptions about relative effect measures (e.g., proportional hazards) are violated. For instance, even with identical hazard ratios (HR), survival curves may differ significantly, affecting state occupancy estimates in economic evaluations and long-term survival extrapolation (Figure 1b).

- ***Sheffield RECReATE project (Researching the use of England's Cancer Registry data for Assessing Treatment Effectiveness) [16-18]***
  - A series of TTE benchmarking studies was developed to assess the feasibility of using real-world data (RWD) to produce reliable clinical evidence. These studies examine whether real-world (English) cancer data is of sufficient quality to generate the RWE to support HTA (Table 2).
  - An extended RCT-RWE four-criteria agreement assessment matrix was developed to complement existing matrices for HTA (Table 3).

Table 2. TTE benchmarking trials in the Sheffield RECReATE project [16-18]

Cancer Type	Cancer Trial Selected for Emulation†	Details
Pancreatic	ESPAC-4	Gemcitabine vs. Gem + Capecitabine
Pancreatic	ACCORD7	FOLFIRINOX vs. Gemcitabine
Pancreatic	CRUK-GEM-CAP	Gemcitabine vs. Gem + Capecitabine
Pancreatic	MPACT	Gemcitabine vs. Gem + Nab-Paclitaxel
Lung	LUX-Lung	Afatinib vs. Gefitinib
Lung	Keynote-024	Pembrolizumab vs. Chemotherapy
Breast	TNT	Carboplatin vs. Docetaxel
Prostate	GUTG-001	Enzalutamide and Abiraterone Sequencing Study, using Flatiron and NCRAS data
Renal Cell	RECORD-3	Sunitinib followed by Everolimus Sequencing Study

NCRAS: National Cancer Registration and Analysis Service.
†For each trial, refer to study protocols for full references. Note: Limited funding and time may prevent completion of all studies. All studies use English NCRAS data unless stated otherwise.

Extended RCT-RWE Agreement Matrix for HTA

Our initial findings from the Sheffield RECReATE project [19] revealed that omitting the assessment of absolute outcomes can be problematic when assessing agreement between emulated and existing RCTs. When systematic bias exists across treatment arms (e.g., immortal time bias), solely examining the RCT-RWE agreement of relative effects (e.g. HR, RR) can mask incorrect estimates of absolute effect differences in RWE, leading to biased decisions if used in cost-effectiveness models (Figure 1).

Figure 1. Toy examples illustrating issues from omitting the assessment of RCT-RWE agreement on absolute effectiveness, despite concordant relative estimates (hazard ratios, HR) between treatment groups: (a) consistent immortal time bias across treatment arms; (b) discordant survival curves

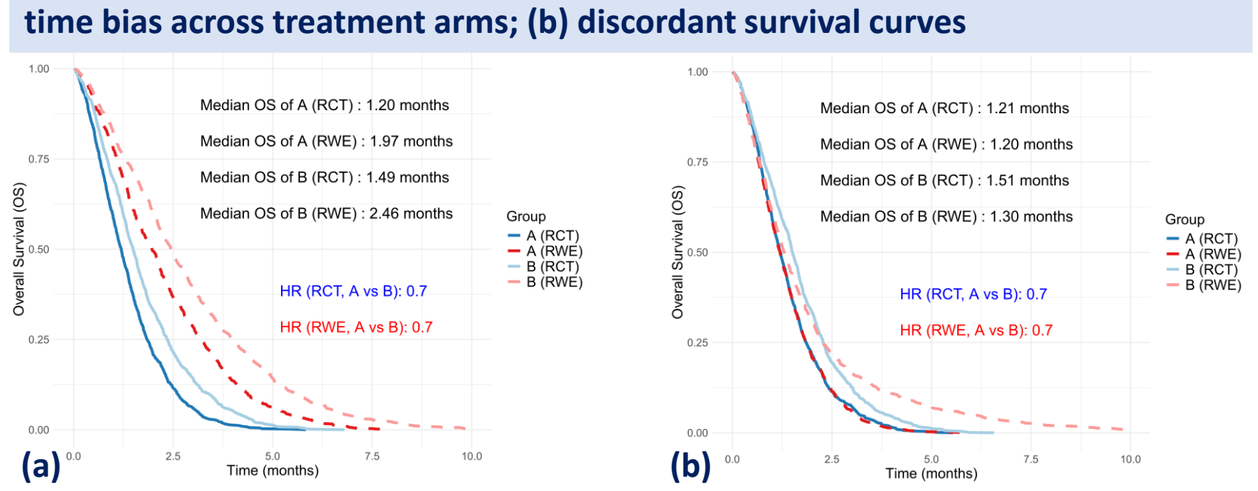

Table 3. Extended RCT-RWE Agreement Assessment Matrix

Criteria	Definition
1. Regulatory agreement	Assess whether the direction and statistical significance of the **comparative effectiveness** (e.g., risk ratio, hazard ratio) from RWE align with the benchmark RCT
2. Estimate agreement	Assessing whether the point estimate of the **comparative effectiveness** from RWE falls within the 95% CIs of those from the benchmark RCT
	Assessing whether the point estimate of the **absolute effectiveness** (e.g., median overall survival) from RWE falls within the 95% CIs of those from the benchmark RCT
3. Exploratory - Standardised difference	$Z = \frac{\hat{\theta}_{RWE} - \hat{\theta}_{RCT}}{\sqrt{\sigma^2_{RWE} + \sigma^2_{RCT}}}$ A Z-value of the **comparative effectiveness** below 1.96 indicates no significant difference between the comparative effectiveness estimates from RWE and benchmark RCT.
4. Exploratory - Survival curve comparison	Assessing whether the point estimates of the RWE **survival curve for each treatment group** fall within the 95% CI of the benchmark trial

CI, confidence interval; RCT, randomised controlled trials, RWE: real-world evidence
- Rows highlighted in **purple** indicate the original three-criteria RCT-RWE Agreement Matrix [1]
- Rows highlighted in **orange** represent the extended criteria we proposed

Conclusion

We propose incorporating extended metrics to assess RCT-RWE agreement (Table 3) in TTE benchmarking studies. This approach provides a more nuanced examination of RWE's reliability and applicability in HTA.

References: [1] Franklin JM, et al. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project. Clin Pharmacol Ther. 2020;107(4):817-826. | [2] Rex S, et al. RWD66 Application of the Target Trial Framework for Estimating Comparative Effectiveness Using Real-World Data-A Systematic Review. Value in Health. 2022; 25(12), S461. | [3] Althunian, T.A., et al., Rivaroxaban was found to be noninferior to warfarin in routine clinical care: A retrospective noninferiority cohort replication study. Pharmacoepidemiology and Drug Safety, 2020. 29(10): p. 1263-1272. | [4] Petito, L.C., et al., Estimates of overall survival in patients with cancer receiving different treatment regimens: emulating hypothetical target trials in the surveillance, epidemiology, and end results (SEER)–Medicare linked database. JAMA network open, 2020. 3(3): p. e200452-e200452. | [5] Yiu, Z.Z., et al., Randomized trial replication using observational data for comparative effectiveness of secukinumab and ustekinumab in psoriasis: a study from the British Association of Dermatologists Biologics and Immunomodulators Register. JAMA dermatology, 2021. 157(1): p. 66-73. | [6] Keyhani, S., et al., Comparative effectiveness of carotid endarterectomy vs initial medical therapy in patients with asymptomatic carotid stenosis. JAMA neurology, 2020. 77(9): p. 1110-1121. | [7] Abrahami, D., et al., Use of Real-World Data to Emulate a Clinical Trial and Support Regulatory Decision Making: Assessing the Impact of Temporality, Comparator Choice, and Method of Adjustment. Clinical Pharmacology & Therapeutics, 2021. 109(2): p. 452-461. | [8] Matthews, A.A., et al., Comparing effect estimates in randomized trials and observational studies from the same population: an application to percutaneous coronary intervention. Journal of the American Heart Association, 2021. 10(11): p. e020357. | [9] Kirchgesner, J., et al., Calibrating Real-World Evidence Studies Against Randomized Trials: Treatment Effectiveness of Infliximab in Crohn's Disease. Clinical Pharmacology & Therapeutics, 2022. 111(1): p. 179-186. | [10] Burn, E., et al., Opioid use, postoperative complications, and implant survival after unicompartmental versus total knee replacement: a population-based network study. The Lancet Rheumatology, 2019. 1(4): p. e229-e236. | [11] Karaboyas, A., et al., Replicating randomized trial results with observational data using the parametric g-formula: an application to intravenous iron treatment in hemodialysis patients. Clinical Epidemiology, 2020: p. 1249-1260. | [12] Admon, A.J., et al., Emulating a novel clinical trial using existing observational data. Predicting results of the PreVent study. Annals of the American Thoracic Society, 2019. 16(8): p. 998-1007. | [13] Lodi, S., et al., Effect estimates in randomized trials and observational studies: comparing apples with apples. American journal of epidemiology, 2019. 188(8): p. 1569-1577. | [14] Bacic, J., et al., Emulating target clinical trials of radical nephrectomy with or without lymph node dissection for renal cell carcinoma. Urology, 2020. 140: p. 98-106. | [15] Boyne, D.J., et al., Association of a shortened duration of adjuvant chemotherapy with overall survival among individuals with stage III colon cancer. JAMA Network Open, 2021. 4(3): p. e213587-e213587. | [16] Latimer NR. Leveraging Real-World Evidence for Health Technology Assessment – Using Big Data to Enable Patient Access: Can we estimate the comparative effectiveness of cancer treatments using UK registry data? ISPOR Europe 2022 Educational Symposia, Nov 6-9, 2022, Vienna. [17] Chang JYA, et al. Leveraging real-world data to assess treatment sequences in health economic evaluations: a study protocol for emulating target trials using the English Cancer Registry and US Electronic Health Records-Derived Database. HEDS Discussion Paper, 2024; 24.01. | [18] Rex S. Protocol for emulating Non-Small Cell Lung and Triple Negative Breast Cancer Target Trials from England's Cancer Registry Data. Report. SCHARR HEDS Discussion Papers (24.02). HEDS Discussion Paper, 2024; 24.02. | [19] Chang JYA. Investigating the Application of Causal Inference Methods for Modelling the Impact of Treatment Sequences in Health Economic Evaluations. 2024 PhD thesis, University of Sheffield.

Jen-Yu Amy Chang was funded by the Wellcome Trust [108903/B/15/Z] and the University of Sheffield. Nicholas Latimer and Saleema Rex received funding from Yorkshire Cancer Research, and Saleema Rex also received funding from Lumanity for this work.

Email: jy.a.chang@sheffield.ac.uk