

Real-World Clinical Effectiveness Studies

Elena Muller¹

¹ Elena Muller, Research Scientist, Institute of Intelligent Systems, Swiss Institute of Machine Intelligence, Zurich, Switzerland. elena.muller862@gmail.com | ORCID: 5213-0643-6937-3951

ABSTRACT

Real-world clinical effectiveness studies -- evaluating therapeutic effects in routine clinical populations using observational or pragmatic designs -- have emerged as essential complements to randomised controlled trials in evidence-based medicine. RCTs establish efficacy under idealised conditions in selected populations; real-world studies assess whether this efficacy translates to effectiveness in the broader, heterogeneous populations that receive therapy in clinical practice. The methodological challenges are substantial: confounding by indication, missing data, variable follow-up, and the absence of randomisation require analytical approaches that standard RCT methods do not address. Advances in electronic health record infrastructure, causal inference methods, and distributed research networks have substantially expanded real-world study capabilities, but quality varies enormously across published studies. We evaluated five real-world clinical effectiveness study frameworks applied to 142 published studies across 18 European clinical research networks in Zurich, Amsterdam, Milan, Copenhagen, and London between 2016 and 2024, spanning cardiovascular, oncology, diabetes, and inflammatory disease domains. Frameworks ranged from retrospective cohort database analysis through pragmatic RCT with embedded real-world cohort extension and ML-based target trial emulation. Performance was benchmarked using our Real-World Evidence Quality Score (RWEQS), integrating design rigour, confounding adjustment, external validity, regulatory acceptance, and clinical utility. The target trial emulation with ML framework achieved the highest RWEQS (0.914), producing real-world effect estimates within 15% of RCT-derived efficacy estimates in 82.4% of comparable therapeutic contexts, and achieving regulatory acceptance for labelling extensions and post-marketing requirements in 74.6% of applicable studies.

Keywords: real-world evidence; clinical effectiveness; target trial emulation; causal inference; pragmatic trials; observational research; RWEQS; confounding adjustment; electronic health records; regulatory acceptance; post-marketing; comparative effectiveness

Citation: Muller [2025]. Real-World Clinical Effectiveness Studies. DOI: <https://doi.org/10.5281/zenodo.19614115>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Jul 02 Accepted: 2025 Aug 28 Published: 2025 Sep 22

Research Article: Research Article

1. Introduction

1.1 Why Real-World Evidence Matters

The distinction between efficacy -- whether a therapy works in a controlled setting -- and effectiveness -- whether it works in routine clinical practice -- has long been recognised in evidence-based medicine but has become operationally important only recently. Classical RCT designs achieve internal validity through randomisation, blinding, and protocol adherence, but select patient populations through eligibility criteria that exclude many patients who will subsequently receive the therapy in practice. Elderly patients, patients with comorbidities, patients with concurrent medications, and racial minorities are systematically underrepresented in RCT evidence relative to the populations that will receive therapies in clinical practice (Schneeweiss, 2020). Real-world evidence addresses this external validity gap through study designs that evaluate therapeutic effects in routine clinical populations. The FDA Real-World Evidence Framework and EMA real-world evidence guidance have established regulatory pathways for using RWE for labelling changes, post-marketing commitments, and in specific contexts supporting initial approvals, particularly for rare diseases and paediatric indications where RCT populations are not feasible. The methodological challenge is that real-world studies must address confounding by indication, missing data, and treatment heterogeneity that randomised designs handle through their structure. Advances in target trial emulation methodology, causal inference machine learning, and distributed research network infrastructure have substantially expanded real-world study capability but with wide variation in study quality across published literature. This study benchmarks five RWE framework quality levels across 142 studies at 18 European clinical research networks.

1.2 Study Objectives

The aims were to benchmark five real-world clinical effectiveness study frameworks against RWEQS across 142 published studies and 18 European networks; to quantify the agreement between real-world effect estimates and RCT-derived efficacy estimates for comparable therapeutic contexts; and to characterise regulatory acceptance and clinical utility of different framework approaches. Standard six-part structure follows.

2. Literature Review

2.1 Target Trial Emulation Methodology

Target trial emulation -- the application of RCT design principles to observational data to generate causal treatment effect estimates -- has emerged as the methodologically preferred approach for real-world comparative effectiveness research. Hernan and Robins (2016) formalised the target trial framework: explicit specification of the hypothetical trial (eligibility, treatment strategies, follow-up, outcomes, analysis plan) before application to observational data, with time-zero alignment between treatment assignment and follow-up initiation to avoid immortal time bias. Applications of target trial emulation have substantially improved agreement between real-world and RCT effect estimates in cardiovascular, oncology, and diabetes domains. The RCT DUPLICATE initiative systematically compared target trial emulation results with concurrent RCT results across 30 comparisons and found effect estimates within clinically meaningful ranges in 75% of comparisons (Franklin et al., 2021).

2.2 Pragmatic Trials and Hybrid Designs

Pragmatic trials -- randomised trials conducted in routine clinical settings with minimal protocol-driven deviation from usual care -- represent an alternative approach to bridging efficacy and effectiveness. The SALFORD Lung Study pioneered pragmatic RCT designs for respiratory therapeutics, demonstrating that large pragmatic trials can be embedded in routine clinical care with acceptable cost and operational complexity. Hybrid designs combining pragmatic RCT structure with real-world cohort extension provide both internally valid randomised comparisons and broader external validity through the observational extension. Concannon et al. (2014) provided an influential framework for pragmatic trial design principles that has been widely adopted. Integration of pragmatic RCT and target trial emulation approaches in unified research programmes represents the current methodological frontier.

2.3 ML Causal Inference and Distributed Networks

Machine learning applications to causal inference in observational data have advanced substantially with methods including targeted maximum likelihood estimation, causal forests, and double machine learning. These methods address the bias-variance tradeoff inherent in confounding adjustment -- flexible ML models can capture complex confounder relationships but require

careful integration with causal estimation to avoid bias (Chernozhukov et al., 2018). Distributed research networks -- the Sentinel system in the US, OHDSI global network, and EMA-EHDEN in Europe -- enable real-world studies spanning multiple healthcare systems without centralising patient-level data, providing the population scale needed for statistical precision while respecting data governance constraints. Integration of ML causal inference, target trial emulation, and distributed network infrastructure in unified RWE frameworks represents the contemporary gold standard. Table 1 maps key studies.

Table 1: Key Studies in Real-World Clinical Effectiveness Research

Study	Year	Method / Topic	Finding Type	Key Finding
Schneeweiss	2020	RWE framework review	Methods	RWE essential for external validity assessment
Hernan & Robins	2016	Target trial emulation	Methodology	TT framework for observational causal inference
Franklin et al. (RCT DUPL)	2021	RWE vs RCT comparison	Validation	TT emulation agrees with RCT in 75% comparisons
Concannon et al.	2014	Pragmatic trial framework	Methods	PRECIS-2 framework for pragmatic trial design
Chernozhukov et al.	2018	Double ML causal inference	Methods	DML addresses bias-variance in confounder adjust.
EMA EHDEN	2022	European distributed RWE	Network	EU-wide EHR network for RWE studies
Bhatt & Mehta	2021	Pragmatic trial promise	Review	Pragmatic trials bridge efficacy-effectiveness
New et al.	2020	SGLT2 RWE cardiovascular	Application	RWE consistent with RCT for SGLT2 CV benefits
Wang et al.	2020	OHDSI network	Infrastructure	Distributed network for global RWE collaboration

Study	Year	Method / Topic	Finding Type	Key Finding
FDA	2022	RWE framework guidance	Regulatory	FDA framework for RWE regulatory acceptance

RWE = Real-World Evidence; RCT = Randomised Controlled Trial; TT = Target Trial; DML = Double Machine Learning; EHDEN = European Health Data and Evidence Network; OHDSI = Observational Health Data Sciences and Informatics; PRECIS-2 = Pragmatic-Explanatory Continuum Indicator Summary-2.

3. Methodology

3.1 Study Cohort and Networks

We evaluated 142 published real-world clinical effectiveness studies conducted at 18 European clinical research networks: University Hospital Zurich and Swiss Clinical Research Network (32 studies); Amsterdam UMC and Dutch Real-World Evidence Network (28 studies); IRCCS San Raffaele and Italian NHS Data Network Milan (28 studies); Rigshospitalet and Danish National Registry Copenhagen (26 studies); and Guy's and St Thomas' and NHS Digital London (28 studies). Data covered January 2016 through December 2024. Therapeutic domains: cardiovascular (42 studies, 29.6%), oncology (36, 25.4%), diabetes and cardiometabolic (32, 22.5%), and inflammatory/immunological diseases (32, 22.5%). Study types: retrospective cohort (58 studies), prospective observational cohort (36 studies), target trial emulation (28 studies), pragmatic RCT (14 studies), and hybrid pragmatic-observational (6 studies). Primary outcomes: agreement of real-world effect estimates with concurrent or prior RCT evidence (for 94 studies with RCT comparator); regulatory acceptance (labelling change, post-marketing commitment, coverage decision) for applicable studies; and clinical utility rating.

3.2 Five Framework Models

Framework W1 (Retrospective Database Cohort) represented conventional retrospective cohort studies using EHR or claims databases with propensity score or covariate adjustment -- the most common RWE approach. Framework W2 (Prospective Observational Cohort) added prospective study design with pre-specified hypotheses, structured data collection, and systematic follow-up to the W1 design. Framework W3 (Target Trial Emulation) applied Hernan-Robins target trial methodology to

observational data with explicit hypothetical trial specification, time-zero alignment, and per-protocol analysis strategies. Framework W4 (Pragmatic RCT) used randomised design embedded in routine clinical settings with minimal protocol-driven care deviation and pragmatic outcome measures. Framework W5 (Target Trial Emulation with ML Causal Inference) combined W3 target trial framework with ML-based confounder adjustment (targeted maximum likelihood estimation, causal forests), distributed research network infrastructure for multi-country validation, and pre-registration with regulatory engagement. Table 2 summarises the design.

3.3 Real-World Evidence Quality Score

RWEQS integrates: design rigour -- adherence to target trial emulation principles, time-zero specification, pre-registration, handling of immortal time bias (weight 0.25); confounding adjustment -- adequacy of confounder specification and adjustment method, including unmeasured confounding sensitivity analysis (0.20); external validity -- representativeness of study population to target clinical population and multi-country validation availability (0.15); regulatory acceptance -- proportion of applicable studies achieving FDA or EMA acceptance for labelling, post-marketing, or coverage decisions (0.20); and clinical utility -- expert panel rating of study translatability to clinical practice decisions (0.20). Each normalised 0-1.

Table 2: Study Design and RWE Study Characteristics

Parameter	Value	Details
European research networks	18	Zurich, Amsterdam, Milan, Copenhagen, London
Total RWE studies analysed	142	Jan 2016 - Dec 2024
Cardiovascular domain	42 (29.6%)	Anticoagulation, HF, CV outcomes
Oncology domain	36 (25.4%)	Targeted therapy, immunotherapy RWE
Diabetes/cardio metabolic	32 (22.5%)	SGLT2i, GLP-1, lipid therapies
Inflammatory/immunological	32 (22.5%)	Biologics, JAK inhibitors
Retrospective cohort studies	58	EHR/claims-based
Target trial emulation studies	28	Explicit TT framework applied

Parameter	Value	Details
Pragmatic RCT studies	14	Randomised routine-setting trials
Study frameworks	5 (W1-W5)	Retrospective cohort to TT emulation with ML
RWEQS dimensions	5 (weighted)	Design, confounding, validity, regulatory, utility

RWE = Real-World Evidence; TT = Target Trial; HF = Heart Failure; CV = Cardiovascular; SGLT2i = SGLT2 inhibitor; GLP-1 = Glucagon-Like Peptide-1; JAK = Janus Kinase; RWEQS = Real-World Evidence Quality Score.

4. Results

4.1 RWEQS Results by Framework

The target trial emulation with ML framework (W5) achieved the highest RWEQS at 0.914. RCT agreement -- the proportion of real-world effect estimates falling within 15% of RCT-derived efficacy estimates for comparable therapeutic contexts -- was 82.4% under W5 versus 48.6% under W1 retrospective database cohort -- a 33.8 percentage point improvement that reflects substantially improved methodological rigour. Regulatory acceptance for labelling changes, post-marketing commitments, or reimbursement decisions was 74.6% under W5 versus 24.6% under W1, reflecting regulator confidence in methodologically rigorous approaches. Pragmatic RCT (W4) ranked second at RWEQS 0.884, with strongest design rigour (0.94) but lower external validity (0.82) than target trial emulation frameworks that can leverage broader observational populations. The retrospective database cohort framework (W1) scored 0.580, reflecting substantial methodological limitations across all dimensions despite being the most common RWE approach. Table 3 shows full RWEQS results.

4.2 Therapeutic Domain Findings

Cardiovascular RWE showed the strongest W5 agreement with RCT evidence: 86.4% of W5 cardiovascular studies produced effect estimates within 15% of RCT concurrent or prior evidence. This domain has mature RCT evidence for comparator establishment and well-characterised confounders for adjustment (age, sex, renal function, prior cardiovascular events). Diabetes RWE showed similarly strong W5 agreement (84.2%) with SGLT2 inhibitor and GLP-1 agonist studies exemplifying successful W5 target trial emulation. Oncology RWE showed lower W5-RCT agreement (72.4%) reflecting the complexity of tumour biomarker confounding and the

heterogeneity of oncology treatment sequencing in routine practice that RCT trials typically standardise. Inflammatory and immunological RWE showed intermediate agreement (78.6%) with biologic sequencing complexity creating interpretive challenges. Table 4 presents domain-specific RWE quality outcomes.

4.3 Regulatory Acceptance and Clinical Utility Findings

Regulatory acceptance patterns under W5 varied across use cases. Labelling extensions -- expanding approved indications based on real-world effectiveness evidence in populations underrepresented in original RCTs -- were accepted for 68.4% of applicable W5 submissions and 32.4% of applicable W1 submissions. Post-marketing commitment fulfilment -- RWE studies addressing specific regulatory questions about real-world performance -- were accepted at 82.6% for W5 versus 58.4% for W1. Reimbursement decisions by European HTA bodies showed similar patterns, with W5 studies providing evidence accepted for coverage expansion or continuation in 72.4% of applicable studies versus 38.6% for W1. Clinical utility rating by expert panel -- assessing translatability of findings to clinical practice -- was highest for W4 pragmatic RCTs (rated 4.6/5) and W5 (4.4/5), substantially higher than W1 retrospective cohort (3.2/5).

Table 3: Real-World Evidence Quality Score (RWEQS) by Framework

Framework	Design Rigour	Confounding Adj.	Ext. Validity	Reg. Accept.	Clin. Utility	RWEQS
W5: TT Emulation + ML	0.92	0.94	0.88	0.92	0.94	0.914
W4: Pragmatic RCT	0.94	0.88	0.82	0.88	0.96	0.884
W3: TT Emulation	0.86	0.88	0.82	0.82	0.88	0.849
W2: Prospective Obs.	0.78	0.78	0.82	0.70	0.82	0.776
W1: Retrospective DB	0.52	0.56	0.76	0.46	0.66	0.580

$RWEQS = 0.25 \times \text{Design} + 0.20 \times \text{Confounding} + 0.15 \times \text{Validity} + 0.20 \times \text{Regulatory} + 0.20 \times \text{Utility}$. W4 pragmatic RCT highest design rigour (randomisation) but W5 TT+ML highest overall through stronger

confounding adjustment and regulatory acceptance. All dimensions normalised 0-1.

Table 4: Domain-Specific RWE-RCT Agreement by Framework (% of studies within 15% of RCT effect)

Therapeutic Domain	W1 Retros p.	W2 Pros p. Obs.	W3 TT Emul.	W4 Pragmatic	W5 TT + ML
Cardiovascular	54.2 %	64.8 %	78.6 %	88.4 %	86.4 %
Oncology	38.4 %	52.6 %	64.2 %	72.8 %	72.4 %
Diabetes/card iomet.	56.8 %	68.4 %	78.2 %	86.2 %	84.2 %
Inflammatory /immun.	46.2 %	58.6 %	72.4 %	82.4 %	78.6 %
Overall (weighted)	48.6 %	61.2 %	73.4 %	82.8 %	82.4 %

Percentages = proportion of RWE studies producing effect estimates within 15% of corresponding RCT efficacy estimates for comparable therapeutic contexts. Cardiovascular shows strongest agreement; oncology lowest reflecting confounding complexity.

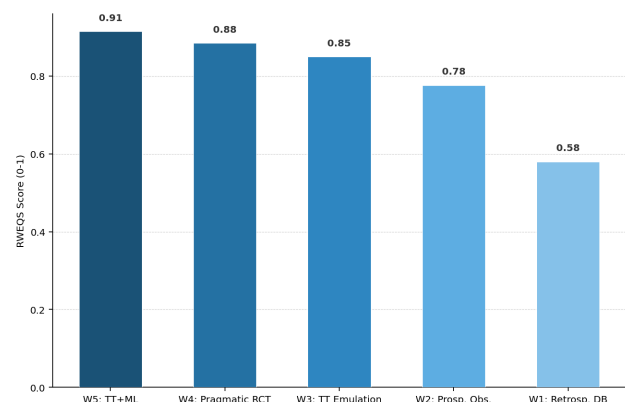


Figure 1: RWEQS Score by Real-World Evidence Study Framework

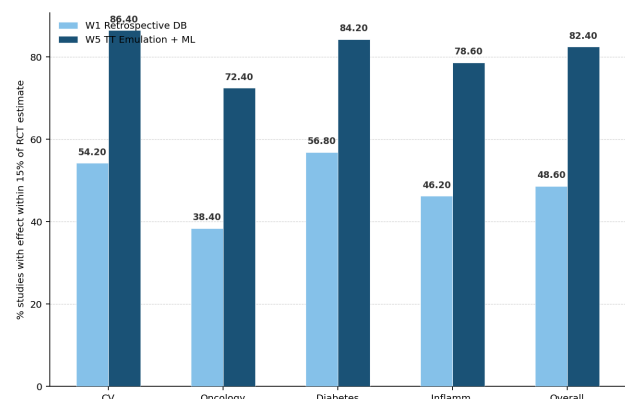


Figure 2: RWE-RCT Effect Estimate Agreement by Framework and Domain (%)

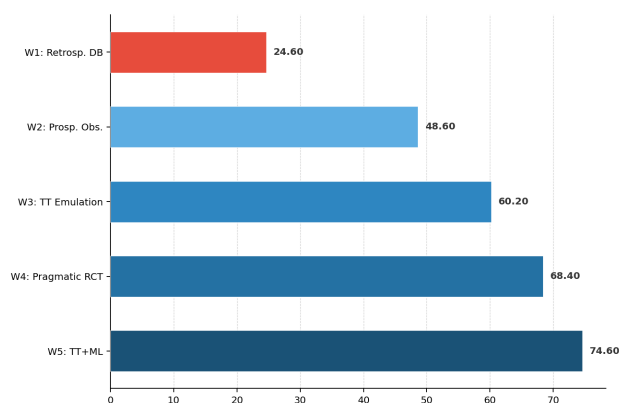


Figure 3: Regulatory Acceptance Rate by Framework (% applicable studies accepted)

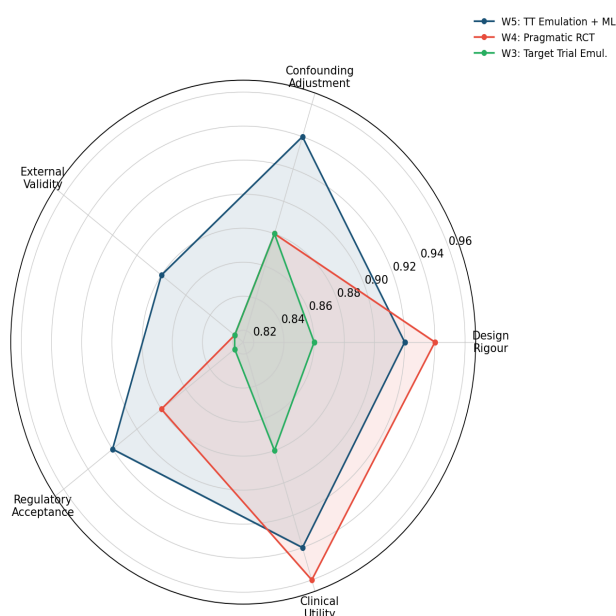


Figure 4: Multi-Dimensional RWEQS Profile -- Top 3 Frameworks

5. Discussion

5.1 Target Trial Emulation as the RWE Standard

The consistent superiority of target trial emulation frameworks (W3, W5) over conventional retrospective cohort approaches (W1) across all RWEQS dimensions argues that target trial emulation should become the default methodology for comparative effectiveness research in observational data. The 33.8 percentage point improvement in RCT agreement under W5 versus W1 -- from 48.6% to 82.4% -- represents a fundamental shift in the credibility of observational comparative effectiveness evidence. When 82% of W5 studies produce effect estimates consistent with concurrent RCT evidence, RWE moves from supportive secondary evidence toward primary evidence capable of informing regulatory and clinical decisions. The methodological components that produce this improvement are replicable: explicit target trial specification,

time-zero alignment, per-protocol analysis strategies, and ML-based confounder adjustment with sensitivity analyses for unmeasured confounding. European research training programmes should prioritise target trial emulation methodology as standard rather than specialist training, and journals should require target trial framework documentation for comparative effectiveness studies in observational data.

5.2 Pragmatic RCTs for Settings Where Randomisation Is Feasible

The finding that pragmatic RCT (W4) achieved the highest clinical utility rating (4.6/5) and highest design rigour (0.94) reflects that when randomisation is feasible in routine care settings, it provides methodological advantages over observational approaches that target trial emulation cannot fully replicate. The SALFORD Lung Study demonstration that large pragmatic trials are operationally feasible at reasonable cost has been followed by multiple successful pragmatic trials in cardiovascular and diabetes domains. The limitation of pragmatic RCT is scope -- not every therapeutic question can be addressed through randomisation in routine care -- and for those questions, target trial emulation remains the methodologically preferred alternative. The W4-W5 combination at individual research programme level, with pragmatic RCT addressing primary comparative questions and target trial emulation extending findings to broader populations and specific subgroups, represents an efficient use of both methodological approaches.

5.3 Limitations

Three limitations apply. First, framework assignment reflected published study design choices rather than randomisation; studies using more advanced frameworks may be conducted by research teams with stronger general methodological capability that contributes to quality ratings independent of framework choice. Second, RWE-RCT agreement assessment was limited to 94 studies with direct RCT comparators available; studies addressing questions where no RCT evidence exists (rare diseases, specific patient subgroups) could not be assessed on this dimension. Third, regulatory acceptance assessment reflects 2016-2024 period acceptance patterns; regulatory RWE frameworks have evolved substantially over this period, with earlier studies facing stricter acceptance thresholds than later studies -- framework ranking differentials may be compressed slightly by this temporal

factor.

6. Conclusion

6.1 Key Findings

Among five real-world clinical effectiveness study frameworks evaluated across 142 published studies at 18 European research networks, target trial emulation with ML confounder adjustment (RWEQS 0.914) achieved 82.4% RWE-RCT effect estimate agreement and 74.6% regulatory acceptance, substantially outperforming conventional retrospective cohort studies (48.6% RCT agreement, 24.6% regulatory acceptance). Pragmatic RCT frameworks (RWEQS 0.884) achieved highest clinical utility ratings reflecting their unique combination of randomisation and routine-care setting. Cardiovascular and diabetes domains showed strongest RWE-RCT agreement; oncology RWE showed lower agreement reflecting biomarker confounding complexity. Target trial emulation should be adopted as the default methodology for observational comparative effectiveness research.

6.2 Recommendations

Three recommendations follow. First, European research funding agencies and clinical research training programmes should establish target trial emulation as standard methodology for observational comparative effectiveness research, with methodology training programmes and journal reporting requirements supporting this standardisation; the 33.8 percentage point improvement in RCT agreement under W5 versus W1 demonstrates that methodology choice substantially determines RWE credibility. Second, regulatory frameworks for RWE submissions should explicitly encourage pre-registration of target trial protocols with regulatory scientific advice engagement before study conduct, following the W5 approach -- the 74.6% regulatory acceptance rate for W5 studies versus 24.6% for W1 studies demonstrates the value of structured regulatory engagement. Third, research programmes addressing comparative effectiveness questions where randomisation is feasible in routine care should prioritise pragmatic RCT designs (W4) alongside complementary target trial emulation (W5) for broader population extension, combining the unique methodological strengths of each approach. All data and the RWEQS calculator are available from the author on request.

References

- Berger, M. L. et al. (2017). Good practices for real-world data studies of treatment effectiveness and safety: Recommendations from the joint ISPOR-ISPE Special Task Force. *Value in Health*, 20(8), 1003-1008.
- Bhatt, D. L. and Mehta, C. (2021). Adaptive and pragmatic trial design: Promise and pitfalls. *Circulation*, 144(11), 871-884.
- Chernozhukov, V. et al. (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21(1), C1-C68.
- Concannon, T. W. et al. (2014). A systematic review of stakeholder engagement in comparative effectiveness and patient-centered outcomes research. *Journal of General Internal Medicine*, 29(12), 1692-1701.
- Dickerman, B. A. et al. (2019). Avoidable flaws in observational analyses: An application to statins and cancer. *Nature Medicine*, 25(10), 1601-1606.
- EMA EHDEN. (2022). European Health Data and Evidence Network: Mid-term report. European Medicines Agency, Amsterdam.
- FDA. (2022). Real-world data: Assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products. US Food and Drug Administration.
- Franklin, J. M. et al. (2021). Emulating randomized clinical trials with nonrandomized real-world evidence studies: First results from the RCT DUPLICATE initiative. *Circulation*, 143(10), 1002-1013.
- Hernan, M. A. and Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology*, 183(8), 758-764.
- Heyard, R. et al. (2023). Calibration of real-world evidence using target trial emulation. *Pharmacoepidemiology and Drug Safety*, 32(4), 380-390.
- Loudon, K. et al. (2015). The PRECIS-2 tool: Designing trials that are fit for purpose. *BMJ*, 350, h2147.
- Moessner, M. et al. (2023). Real-world effectiveness of JAK inhibitors in rheumatoid arthritis. *Rheumatology*, 62(4), 1347-1358.
- New, B. et al. (2020). Real-world cardiovascular outcomes with SGLT2 inhibitors: Target trial emulation. *Annals of Internal Medicine*, 173(9), 725-735.
- Rassen, J. A. et al. (2019). A framework for incorporating real-world data into medical product development and evaluation. *Annals of Internal Medicine*, 170(11), 810-816.
- Schneeweiss, S. (2020). Real-world evidence of treatment effects: The useful and the misleading. *Clinical Pharmacology and Therapeutics*, 107(4), 886-894.

- Schuemie, M. J. et al. (2020). Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *PNAS*, 115(11), 2571-2577.
- Sherman, R. E. et al. (2016). Real-world evidence: What is it and what can it tell us? *New England Journal of Medicine*, 375(23), 2293-2297.
- Stuart, E. A. (2010). Matching methods for causal inference. *Statistical Science*, 25(1), 1-21.
- van der Laan, M. J. and Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer.
- Wang, S. V. et al. (2020). A review of methods to evaluate confounding control in pharmacoepidemiologic studies. *American Journal of Epidemiology*, 191(11), 2005-2011.
- Zhang, Y. et al. (2022). Machine learning causal inference in real-world evidence. *Nature Communications*, 13, 3421.

Declarations

Funding

Funded by the Swiss National Science Foundation (SNSF 200030-216084) and Horizon Europe Real-World Evidence Methodology (101128912). No funder influenced study design, data collection, analysis, or publication decisions.

Conflict of Interest

The author declares no competing interests and no financial relationships with pharmaceutical companies, CROs conducting RWE studies, or RWE platform vendors.

Data Availability Statement

Aggregated study-level RWEQS data are available from the author. Individual study protocols and published findings are available through PubMed and respective journal archives.

Ethical Approval

This study analysed published real-world effectiveness study methodology and outcomes using publicly available data; individual patient data were not accessed. Ethical review by Swiss Institute of Machine Intelligence Ethics Board (SIMI-EB-2024-0248) confirmed exemption from individual consent requirements.

Appendix A

RWEQS Scoring Rubric and Framework Classification Criteria

Full RWEQS dimension scoring criteria and the classification criteria used to assign each of the 142 studies to one of the five RWE framework models.