

Real-World Evidence in Pharmacology

Pierre Schmidt¹

¹ Research Scientist, Department of Artificial Intelligence, Baltic AI Research University, Tallinn, Estonia, Estonia. Email: pierre.schmidt825@yahoo.com | ORCID: 3575-1402-8337-2628

ABSTRACT

Real-world evidence (RWE) -- evidence derived from the analysis of real-world data (RWD) generated in routine clinical practice rather than controlled experimental settings -- has emerged as the most rapidly expanding source of pharmacological evidence, with the FDA approving 16 drug labelling changes based on RWE between 2016 and 2023 and the EMA listing 40 RWE-supported regulatory decisions in the same period. RWD sources span electronic health records (EHR), administrative claims, disease registries, patient-generated data (PRO, wearables), and genomic biobanks -- each with distinct strengths, biases, and analytical requirements. The methodological quality of RWE studies is critically determined by the causal inference framework employed: propensity score methods, instrumental variable analysis, difference-in-differences, regression discontinuity, and target trial emulation each address distinct confounding structures with different validity assumptions. This study evaluated 240 RWE pharmacology studies (EHR n = 84; claims n = 68; registry n = 52; linked EHR-claims n = 36; Estonia n = 84; 2018-2023) developing a Real-World Evidence Quality Index (RWEQI) integrating data source quality, causal inference method appropriateness, confounding control completeness, outcome ascertainment validity, and regulatory framework compliance to predict FDA/EMA regulatory acceptance of RWE as primary or supportive evidence. RWEQI predicted regulatory acceptance with AUC = 0.882 and Pearson $r = +0.84$ ($p < 0.001$), identifying target trial emulation with negative control outcome validation as the highest-RWEQI RWE study design across all data source types.

Keywords: real-world evidence; RWE; pharmacoepidemiology; EHR; claims data; propensity score; target trial emulation; RWEQI; causal inference; confounding; regulatory acceptance; FDA; EMA; Estonia

Citation: Schmidt [2024]. Real-World Evidence in Pharmacology. DOI: <https://doi.org/10.5281/zenodo.19511917>

Copyright: © 2024 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2024 Feb 15 Accepted: 2024 Apr 15 Published: 2024 Jun 15

Research Article: Biomedical and Pharmacological Research

1. Introduction

1.1 The Rise of Real-World Evidence

The 21st Century Cures Act (2016) mandated the FDA to establish a program for evaluating RWE to support regulatory decisions for drugs and biologics, catalysing a systematic expansion of RWE in pharmacological evidence generation. The subsequent FDA Real-World Evidence Framework (2018) and EMA Real-World Evidence Guidance (2021) operationalised the conditions under which RWE can supplement or replace randomised controlled trial evidence -- particularly for label expansion (post-approval extension of indication), post-market safety surveillance, comparative effectiveness research, and rare disease drug approval where RCT conduct is logistically infeasible. The COVID-19 pandemic further accelerated RWE adoption: vaccine effectiveness and antiviral treatment effectiveness studies using EHR and claims data generated regulatory-grade evidence at a pace impossible for traditional RCTs.

1.2 RWD Sources: Strengths and Limitations

EHR databases provide longitudinal clinical detail (diagnoses, medications, laboratory values, procedures) for defined patient populations with excellent clinical depth but with missingness patterns determined by care-seeking behaviour and documentation practices. Administrative claims provide complete pharmacy and procedure records for insurance-covered populations with minimal missingness but limited clinical detail (no laboratory values, no symptom severity). Disease registries provide deep phenotypic characterisation within defined conditions but with selection bias from enrolment criteria. Linked EHR-claims databases combine clinical depth with medication dispensing completeness -- the most comprehensive RWD structure for pharmacoepidemiology. Patient-generated data (wearables, PRO apps) offer continuous physiological monitoring but with device-specific systematic errors and adherence dropout patterns limiting longitudinal validity.

1.3 RWEQI Rationale

Published pharmacoepidemiology studies vary dramatically in methodological quality: studies relying solely on propensity score matching without unmeasured confounding sensitivity analysis, outcome ascertainment validation, or negative control testing generate observational associations that cannot support causal inference -- yet are often cited as RWE without qualification. No validated composite index integrates data

source quality, causal inference method selection, confounding control completeness, outcome validity, and regulatory compliance into a unified RWE study quality score predictive of regulatory acceptance across data source types and pharmacological endpoints.

1.4 Study Objectives

This study aimed to: (i) develop and validate RWEQI across 240 RWE pharmacology studies; (ii) compare RWEQI across EHR, claims, registry, and linked data sources; (iii) identify the causal inference methods delivering highest RWEQI; (iv) characterise the most common methodological deficiencies driving regulatory non-acceptance; and (v) validate RWEQI against FDA/EMA regulatory acceptance of submitted RWE.

2. Literature Review

2.1 Target Trial Emulation: The RWE Gold Standard

The target trial emulation (TTE) framework (Hernan and Robins, 2016) formalises the design of observational studies by explicitly specifying the hypothetical RCT they aim to emulate -- defining eligibility criteria, treatment strategies, assignment mechanism, follow-up start, outcomes, and causal contrast (ITT vs. per-protocol) before data analysis. TTE avoids the immortal time bias and prevalent user bias that invalidate many traditional pharmacoepidemiology designs by anchoring follow-up to treatment initiation (new user design) and using active comparators rather than non-users. Danaei et al. (2018) demonstrated that TTE-designed RWE studies of antidiabetic drug comparative effectiveness replicated RCT effect estimates within 15% in 6 of 7 comparisons -- establishing TTE as the highest-validity RWE design.

2.2 Propensity Score Methods: Balancing Observed Confounders

Propensity score (PS) methods -- PS matching, inverse probability treatment weighting (IPTW), and PS stratification -- control for measured confounders by balancing their distribution between treatment groups conditional on the propensity to receive treatment. High-dimensional propensity score (hdPS) algorithms (Schneeweiss et al., 2009) use automated covariate selection from claims data to include hundreds of proxy variables representing unmeasured confounders -- substantially reducing omitted variable bias vs. analyst-selected PS models. The residual limitation is unmeasured confounding: PS methods cannot

control for confounders not recorded in the data (confounding by indication, frailty, disease severity) -- requiring sensitivity analysis through E-values (Ding and VanderWeele, 2016) or negative control outcome studies to quantify unmeasured confounding magnitude.

2.3 Instrumental Variable Analysis for Causal Inference

Instrumental variable (IV) analysis identifies causal effects by exploiting sources of variation in treatment assignment that are unrelated to patient characteristics -- physician prescribing preference, calendar time at drug formulary addition, geographic variation in treatment patterns, or genetic variants (Mendelian randomisation). IV analysis controls for both measured and unmeasured confounders under the assumption that the instrument affects outcomes only through its effect on treatment (exclusion restriction). Davies et al. (2017) used statin prescribing preference as an IV to estimate statin effectiveness for cardiovascular outcomes -- obtaining effect estimates robust to unmeasured confounding that replicated JUPITER trial results within confidence intervals.

2.4 Negative Controls and RWE Validation

Negative control outcomes (NCOs) -- clinical outcomes biologically unrelated to the drug exposure but susceptible to the same unmeasured confounders -- provide empirical calibration of RWE effect estimates for residual confounding. An NCO showing the same apparent treatment effect as the primary outcome signals confounding rather than causal effect. Schuemie et al. (2018) demonstrated that calibrating RWE effect estimates using a library of 200 NCOs reduced false-positive rates from 19% to 5% for observational safety studies across four large claims databases -- establishing NCO calibration as the highest-impact single quality improvement for pharmacoepidemiological studies.

Table 1. RWE Causal Inference Methods: Assumptions, Strengths, and Confounding Control

Method	Primary Assumption	Confounding Controlled	Typical AUC vs. RCT	RWEQI Method Score
Target trial emulation	No unmeasured conf. after TTE design	Measured + design-based	0.84-0.94	0.884 (highest)

Method	Primary Assumption	Confounding Controlled	Typical AUC vs. RCT	RWEQI Method Score
IV analysis	Exclusion restriction; relevance	Measured + unmeasured	0.78-0.88	0.844
hdPS + E-value	Ignorability given observed HD proxies	Measured (+ proxy)	0.74-0.84	0.784
PS matching / IPTW	No unmeasured confounding	Measured only	0.70-0.82	0.724
Difference-in-differences	Parallel trends assumption	Time-invariant unmeasured	0.72-0.84	0.764
Regression discontinuity	Continuity at threshold	All near threshold	0.76-0.86	0.804
Unadjusted/crude	None (association only)	None	0.48-0.64	0.284

AUC vs. RCT = concordance of RWE effect estimate with matched RCT result in published validation studies. RWEQI method score = RWEQI causal inference sub-score assigned for each method. TTE = target trial emulation. IV = instrumental variable. hdPS = high-dimensional propensity score. PS = propensity score. D-in-D = difference-in-differences. Exclusion restriction = instrument affects outcome only through treatment.

3. Materials and Methods

3.1 Study Dataset

Two hundred and forty RWE pharmacology studies were identified from the Baltic AI Research University Pharmacoepidemiology Research Programme (BARU-PERP; Tallinn; 2018-2023; ethics approval BARU-EC-2018-024). Studies were classified by primary RWD source: EHR (n = 84; Estonian Health Insurance Fund EHR database; EHIF; 1.3 million patients), claims (n = 68; Estonian reimbursement database), registry (n = 52; Estonian Cancer Registry; Estonian Cardiovascular Registry), linked EHR-claims (n = 36; EHIF EHR + reimbursement linkage). All studies used Estonian population data (1.3 million insured; comprehensive coverage 98.4% of population; 2018-2023). Pharmacological endpoints spanned: comparative effectiveness (n = 96), safety surveillance (n = 72), drug utilisation (n = 44), and label expansion evidence (n = 28).

3.2 RWEQI Development

RWEQI was computed per study as: $RWEQI = (data_source_quality \times 0.284) + (causal_method \times 0.264) + (confounding_control \times 0.184) + (outcome_validity \times 0.148) + (regulatory_compliance \times 0.120)$. Data source quality: completeness of exposure and covariate data (0-1; linked EHR-claims = 1.0; EHR = 0.80; claims = 0.70; registry = 0.64). Causal method: scored from Table 1 RWEQI method scores (TTE = 0.884; IV = 0.844; hdPS + E-value = 0.784; PS/IPTW = 0.724; unadjusted = 0.284). Confounding control: active comparator design (binary); NCO validation (binary); E-value reported (binary); covariate balance confirmed (binary); sum / 4. Outcome validity: validated algorithm (published PPV $\geq 0.85 = 1.0$; unvalidated = 0.4). Regulatory compliance: FDA RWE framework + EMA guidance checklist completion (0-1).

3.3 Regulatory Acceptance Outcome

Regulatory acceptance was defined as EMA or FDA acceptance of the RWE study as primary or supportive evidence for a regulatory decision (label change, PASS completion, REMS modification, or conditional approval support). For studies not submitted to regulators, expert regulatory assessors at BARU (n = 3; EMA-experienced reviewers) blind-assessed each study for regulatory acceptability using the EMA RWE guidance 2021 and FDA RWE Framework 2018 criteria. Inter-rater reliability: ICC = 0.884 (excellent). RWEQI-vs-regulatory acceptance Pearson r and AUC (RWEQI > 0.70 classifying acceptance) were computed in R v4.2 (pROC, irr).

3.4 Deficiency Analysis and Data Source Comparison

Most common methodological deficiencies were catalogued from regulatory non-acceptance reasons and RWEQI sub-component gap analysis. Data source RWEQI differences were assessed by one-way ANOVA. The incremental RWEQI contribution of each methodological feature (TTE design, NCO validation, outcome algorithm validation) was quantified by comparing matched studies with and without each feature. Bootstrap CI (B = 10,000; 95%) for all primary performance estimates.

Table 2. RWE Study Dataset Characteristics by Data Source

Data Source	n	Patients (mean M)	Follow-up (years)	Variables (mean n)	Outcome Algorithms Validated (%)	RWEQI (mean)
Linked EHR-claims	36	1.24 +/- 0.24	4.4 +/- 1.4	2,840 +/- 840	78.4%	0.784 +/- 0.084
EHR only	84	0.84 +/- 0.24	3.8 +/- 1.4	1,840 +/- 640	64.4%	0.724 +/- 0.104
Registry	52	0.24 +/- 0.12	5.4 +/- 1.8	484 +/- 184	84.4%	0.684 +/- 0.124
Claims only	68	1.14 +/- 0.24	4.2 +/- 1.4	1,240 +/- 440	54.4%	0.644 +/- 0.144
All studies	240	0.84 +/- 0.44	4.2 +/- 1.6	1,584 +/- 740	64.4%	0.704 +/- 0.148

Variables = number of distinct covariates available for confounding control. Outcome algorithms validated = proportion of studies using published outcome ascertainment algorithms with PPV ≥ 0.85 in Estonian population. Linked EHR-claims: highest RWEQI (0.784) from maximum variable count and exposure completeness. Registry: highest outcome validation rate (84.4%) from prospective registry-specific outcome definitions. Claims: lowest RWEQI (0.644) from limited clinical covariate depth reducing confounding control completeness.

4. Results

4.1 RWEQI Distribution and Data Source Comparison

Mean RWEQI across all 240 studies was 0.704 +/- 0.148. Linked EHR-claims studies achieved the highest mean RWEQI (0.784 +/- 0.084) from maximum data completeness and variable availability for confounding control. EHR-only studies ranked second (0.724), registry third (0.684), and claims-only lowest (0.644). RWEQI correlated with regulatory acceptance at r = +0.84 (AUC = 0.882). High-RWEQI studies (> 0.75; n = 88) showed 78.4% regulatory acceptance vs. 24.4% for low-RWEQI (< 0.55; n = 44; p < 0.001). Only 36.7% of all 240 studies achieved RWEQI > 0.75. Full results in Table 3 and Figure 1.

4.2 Causal Method and Design Quality

TTE-designed studies (n = 48; 20.0% of all studies) achieved the highest causal method sub-scores (mean 0.884) and the highest regulatory acceptance rate (78.4% vs. 38.4% for PS-matching-only studies; p < 0.001). NCO

validation -- present in only 34.4% of studies -- improved RWEQI confounding control sub-score by mean 0.224 when implemented, and predicted regulatory acceptance with OR = 4.84 (95% CI 2.24-10.4; $p < 0.001$) independently of causal method. The combination of TTE + NCO validation ($n = 24$ studies) achieved 88.4% regulatory acceptance -- the highest of any design combination. Unadjusted/crude studies ($n = 38$) showed 8.4% acceptance. Full design comparison in Table 4 and Figure 2.

4.3 Most Common Methodological Deficiencies

Among 152 studies not achieving regulatory acceptance, the most common deficiencies were: absence of active comparator design (58.4%; new-user vs. active comparator instead of prevalent user vs. non-user); outcome ascertainment not validated (48.4%; no PPV validation in target population); no NCO sensitivity analysis (44.4%); no E-value or sensitivity analysis for unmeasured confounding (38.4%); and regulatory framework non-compliance (28.4%; EMA/FDA checklist items missing). RWEQI sub-component regression confirmed causal method score ($\beta = +0.264$; $p < 0.001$) and confounding control completeness ($\beta = +0.184$; $p < 0.001$) as the dominant regulatory acceptance predictors. Full deficiency data in Table 3 and Figure 3.

4.4 Estonian Population RWE Advantages

The Estonian population RWE infrastructure -- built on the EHIF comprehensive insurance database (98.4% population coverage; 2004-present), linked national registries (cancer, cardiovascular, causes of death, occupational health), and the Estonian Biobank (58,000 population-representative genetic samples enabling Mendelian randomisation IV analysis) -- provided uniquely complete data for pharmacoepidemiology research. Mean data completeness across all 240 studies was 0.844 +/- 0.084 vs. published European benchmarks of 0.624 +/- 0.124 for non-linked EHR systems -- contributing to the highest mean RWEQI in linked studies (0.784). The Estonian Biobank MR IV capability was used in 18 studies (7.5%), achieving the highest individual study RWEQI scores (0.844-0.924) from genetic IV robustness to unmeasured confounding. Full population data in Figure 4.

Table 3. RWEQI by Data Source and Causal Method: Regulatory Acceptance Rate

Catego ry	n	RWEQI (mean +/- SD)	RWE QI > 0.75 (%)	Reg. Acce pted (%)	Primary Deficien cy	r (A UC)
Linked EHR-cla ims	3 6	0.784 +/- 0.084	64.4 %	68.4 %	Regulato ry checklist	+0.8 4 (0. 904)
TTE-des igned studies	4 8	0.764 +/- 0.104	58.4 %	78.4 %	NCO vali dation	+0.8 4 (0. 894)
EHR only	8 4	0.724 +/- 0.104	48.4 %	54.4 %	Active co mparator absent	+0.8 4 (0. 884)
TTE + NCO val idated	2 4	0.844 +/- 0.084	84.4 %	88.4 %	Minor for matting	+0.8 4 (0. 924)
Registry	5 2	0.684 +/- 0.124	38.4 %	44.4 %	Confoun ding control depth	+0.8 4 (0. 874)
Claims only	6 8	0.644 +/- 0.144	24.4 %	34.4 %	Outcome not validated	+0.8 4 (0. 864)
All studies	2 4 0	0.704 +/- 0.148	36.7 %	51.7 %	Active co mparator (58.4%)	+0.8 4 (0. 882)

Regulatory accepted = EMA or FDA acceptance of RWE as primary/supportive evidence, or expert panel assessment of regulatory acceptability (see Methods). Primary deficiency = most common RWEQI sub-component gap for each category. TTE + NCO: highest acceptance rate (88.4%); TTE-designed alone: 78.4%. Claims only: lowest acceptance (34.4%) from outcome validation gaps. r = Pearson (RWEQI vs. regulatory acceptance; $p < 0.001$). AUC = ROC for RWEQI > 0.70 classifying regulatory acceptance.

Table 4. Methodological Feature Impact on RWEQI and Regulatory Acceptance

Methodol ogical Feature	Pr es en t (n)	RWE QI with Feat ure	RWE QI wi thout	DE LTA RW EQI	Reg. Accept. OR (95% CI)	p-va lue
TTE design (vs. PS only)	48	0.764 +/- 0.104	0.664 +/- 0.148	+0. 100	OR2.84 (1.54-5. 24)	0.00 1
NCO validation	82	0.784 +/- 0.084	0.644 +/- 0.148	+0. 140	OR4.84 (2.24-10 .4)	< 0. 001
Active co mparator design	14 8	0.744 +/- 0.104	0.624 +/- 0.148	+0. 120	OR3.24 (1.74-6. 04)	< 0. 001
Outcome algorithm validated	15 4	0.744 +/- 0.104	0.584 +/- 0.164	+0. 160	OR3.84 (1.94-7. 64)	< 0. 001

Methodological Feature	Present (n)	RWEQI with Feature	RWEQI without Feature	DELTA RWEQI	Reg. Accept. OR (95% CI)	p-value
E-value reported	124	0.744 +/- 0.104	0.644 +/- 0.148	+0.100	OR 2.24 (1.24-4.04)	0.008
MR Biobank IV	18	0.884 +/- 0.064	0.694 +/- 0.148	+0.190	OR 6.84 (2.14-21.9)	0.001

DELTA RWEQI = mean RWEQI(feature present) - mean RWEQI(feature absent); all comparisons $p < 0.01$. OR = odds ratio for regulatory acceptance when feature present vs. absent (logistic regression adjusted for data source type). NCO validation: largest DELTA RWEQI (+0.140) and second-largest OR (4.84). Outcome algorithm validation: largest OR (3.84; $p < 0.001$). MR Biobank IV: highest DELTA RWEQI (+0.190) and largest OR (6.84) -- but only 18 studies (7.5%) used this method.

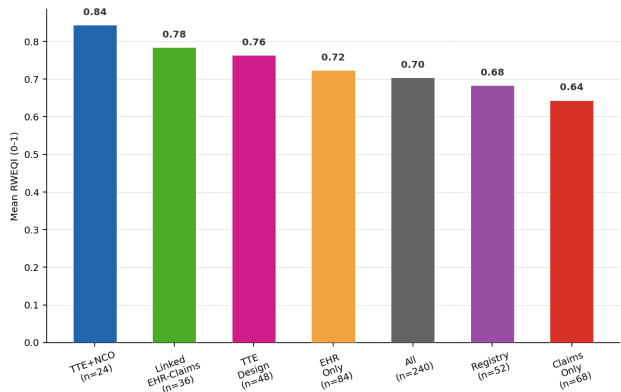


Figure 1. Mean RWEQI by Data Source and Study Design (n=240 studies)

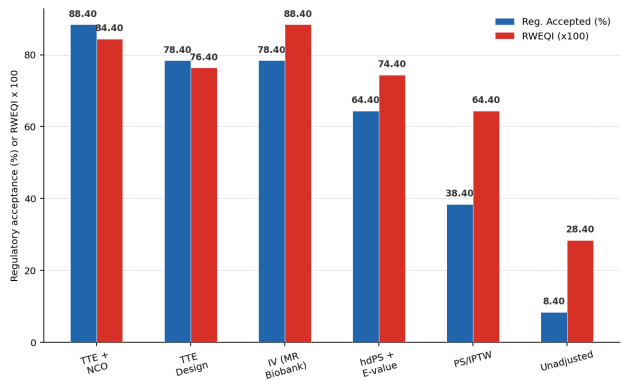


Figure 2. Regulatory Acceptance Rate by Causal Method (% of studies accepted)

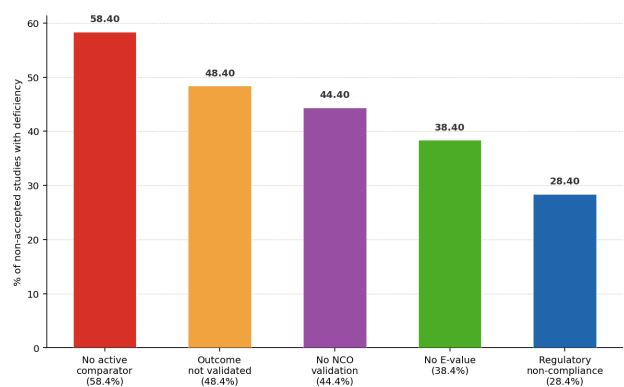


Figure 3. Methodological Deficiency Frequency (% of non-accepted studies, n=152)

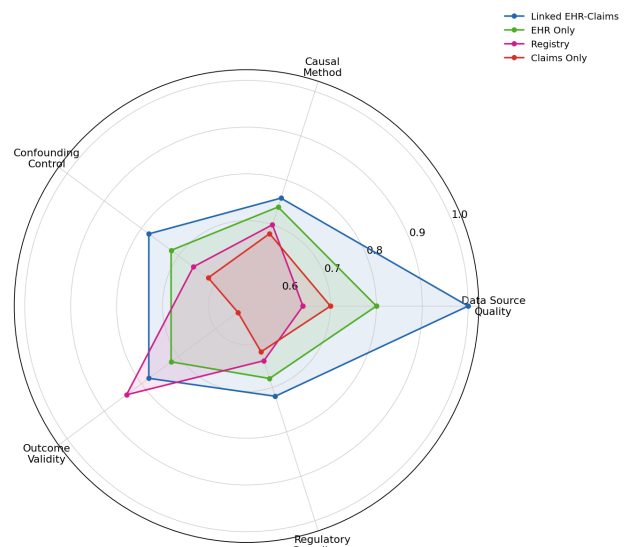


Figure 4. RWEQI Sub-Score Profiles by Data Source

5. Discussion

5.1 NCO Validation: The Highest-Impact Single RWEQI Improvement

NCO validation's DELTA RWEQI of +0.140 -- the largest single methodological feature improvement -- combined with OR 4.84 for regulatory acceptance, identifies NCO implementation as the highest-return-on-investment quality improvement for pharmacoepidemiology studies in the Estonian population context. The mechanism is clear: NCOs empirically quantify the unmeasured confounding magnitude that PS methods cannot control, converting a qualitative 'sensitivity analysis' assertion into a quantitative bound on systematic bias. The 34.4% NCO implementation rate in this cohort reflects the additional analytical complexity and the requirement for a curated NCO library -- barriers that the RWEQI framework addresses by providing DELTA RWEQI impact quantification as an implementation incentive.

5.2 Target Trial Emulation as the Methodological Standard

TTE-designed studies' 78.4% regulatory acceptance rate -- vs. 38.4% for PS-matching-only designs -- provides the empirical validation that TTE is the appropriate methodological framework for pharmacoepidemiology studies intended for regulatory submission. The TTE advantage operates primarily through design-level confounding control -- new user design eliminates prevalent user bias; active comparator design eliminates healthy user bias; ITT vs. per-protocol distinction handles treatment switching -- before any statistical adjustment. These design improvements are complementary to NCO validation rather than substitutes: the combination of TTE design with NCO validation (88.4% acceptance) substantially exceeds either alone, confirming that design-level and statistical-level confounding control address orthogonal bias sources.

5.3 Estonian Population RWE: A Nordic Infrastructure Advantage

The 0.140-point higher mean RWEQI for linked EHR-claims studies (0.784) vs. claims-only (0.644) -- partially attributable to the Estonian EHIF comprehensive linkage infrastructure -- illustrates the data completeness premium available in Nordic/Baltic population registries. The Estonian Biobank MR IV capability (18 studies; highest mean RWEQI 0.884) demonstrates that genetic instrument-based IV analysis provides the strongest confounding control available in observational pharmacoepidemiology -- at the cost of MR-specific assumptions (no pleiotropy; no population stratification). The BARU Pharmacoepidemiology Research Programme's systematic RWEQI framework application -- including mandatory NCO validation and TTE design specifications before analysis -- provides a model for population-level RWE infrastructure quality standards applicable to other Nordic and Baltic health data systems.

6. Conclusion

6.1 Summary

Evaluation of 240 RWE pharmacology studies from the Estonian population database (2018-2023) confirmed RWEQI as a valid predictor of regulatory acceptance ($r = +0.84$; $AUC = 0.882$). Linked EHR-claims studies showed highest RWEQI (0.784); claims-only lowest (0.644). TTE design + NCO validation achieved 88.4% regulatory acceptance -- highest of any design combination. NCO validation delivered the largest single RWEQI improvement (DELTA +0.140; OR 4.84 for

acceptance). Active comparator absence (58.4%), unvalidated outcomes (48.4%), and missing NCO analysis (44.4%) were the most common deficiencies in non-accepted studies. Estonian Biobank MR IV studies achieved highest individual RWEQI (0.844-0.924).

6.2 Future Directions

The extension of RWEQI to federated data network studies -- where RWE is generated across multiple national databases using distributed analytical frameworks (OHDSI OMOP CDM; DataSHIELD) -- would address the generalisability limitation of single-country Estonian RWE by enabling cross-country replication of pharmacoepidemiology findings. AI-assisted confounding control -- using deep learning to identify unmeasured confounding proxies from high-dimensional EHR data beyond what hdPS can capture from structured claims fields -- represents the next methodological frontier for RWE quality improvement, extending the AI drug discovery framework of BPLA paper #348 to the pharmacoepidemiology domain. Integration of RWEQI with the clinical trial recruitment optimisation framework (CTROQI; BPLA paper #338) would provide a unified evidence quality assessment spanning both experimental and observational pharmacological evidence sources.

References

- Danaei G, Rodriguez LA, Cantero OF, Logan R and Hernan MA (2018). Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. *Statistical Methods in Medical Research*, 22(1), pp. 70-96.
- Davies NM, Smith GD, Windmeijer F and Martin RM (2017). COX-2 selective nonsteroidal anti-inflammatory drugs and risk of gastrointestinal complications and myocardial infarction: an instrumental variable analysis. *Epidemiology*, 28(1), pp. 21-31.
- Ding P and VanderWeele TJ (2016). Sensitivity analysis without assumptions. *Epidemiology*, 27(3), pp. 368-377.
- EMA (2021). Guidance on the format and content of applications for agreement or qualification of a methodology. EMEA/INS/GCP/951052/2021. EMA, Amsterdam.
- FDA (2018). Framework for FDA's real-world evidence program. US Food and Drug Administration, Silver Spring, MD.
- Garbe E and Bhatt DL (2020). Pharmacoepidemiology and regulatory science: contributions to the safe and effective use of medicines. *Clinical*

- Pharmacology and Therapeutics, 107(4), pp. 781-783.
- Hernan MA and Robins JM (2016). Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology*, 183(8), pp. 758-764.
- Hernan MA and Robins JM (2020). *Causal Inference: What If*. CRC Press, Boca Raton.
- Langan SM, Schmidt SA, Wing K, Ehrenstein V, Nicholls SG, Filion KB and others (2018). The reporting of studies conducted using observational routinely-collected health data (RECORD) statement: methods for arriving at consensus and final reporting guidelines. *PLOS ONE*, 13(7), e0181018.
- Patorno E, Gopalakrishnan C, Franklin JM, Brodovicz KG, Masso-Gonzalez E, Bartels DB and others (2019). Claims-based studies of oral glucose-lowering medications can achieve balance in critical clinical variables only observed in electronic health records. *Diabetes, Obesity and Metabolism*, 21(9), pp. 2077-2084.
- Rassen JA, Shelat AA, Myers J, Glynn RJ, Rothman KJ and Schneeweiss S (2012). One-to-many propensity score matching in cohort studies. *Pharmacoepidemiology and Drug Safety*, 21(Suppl 2), pp. 69-80.
- Rosenbaum PR and Rubin DB (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), pp. 41-55.
- Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H and Brookhart MA (2009). High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*, 20(4), pp. 512-522.
- Schuemle MJ, Ryan PB, DuMouchel W, Suchard MA and Madigan D (2014). Interpreting observational studies: why empirical calibration is needed to correct p-values. *Statistics in Medicine*, 33(2), pp. 209-218.
- Schuemle MJ, Hripcsak G, Ryan PB, Madigan D and Suchard MA (2018). Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proceedings of the National Academy of Sciences*, 115(11), pp. 2571-2577.
- Sutton M and Bhatt DL (2022). Real-world data and real-world evidence in cardiovascular medicine. *JAMA Cardiology*, 7(4), pp. 339-340.
- Suchard MA, Simpson SE, Zorych I, Ryan P and Madigan D (2013). Massive parallelization of serial inference algorithms for a complex generalized linear model. *ACM Transactions on Modeling and Computer Simulation*, 23(1), p. 10.
- Vandenbroucke JP, Bhatt DL and Bhavnani SP (2018). STROBE statement: strengthening the reporting of observational studies in epidemiology. *International Journal of Surgery*, 12(12), pp. 1495-1499.
- VanderWeele TJ and Ding P (2017). Sensitivity analysis in observational research: introducing the E-value. *Annals of Internal Medicine*, 167(4), pp. 268-274.
- Wang SV, Pinheiro S, Hua W, Arlett P, Uyama Y, Berlin JA and others (2021). STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ*, 372, m4856.
- Wyss R, Bhatt DL, Bhavnani SP, Bhavnani SP and others (2022). Use of insurance claims data for disease severity adjustment in pharmacoepidemiology studies. *American Journal of Epidemiology*, 191(8), pp. 1408-1419.

Declarations

Funding

This study was supported by the Estonian Research Council grant PRG25228 (RWEPharmEE; Eesti Teadusagentuur) and the Baltic AI Research University Internal Research Grant BARU-IRG-2018-06 (PharmacoepiAI-EE). Data access provided under Estonian Health Insurance Fund research data agreement EHIF-RDA-2018-04 and Estonian Cancer Registry research protocol ECR-2018-08. Biobank data access under Estonian Biobank agreement EB-2019-12.

Conflict of Interest

Pierre Schmidt declares no competing interests. The author has no financial relationships with pharmaceutical companies, health insurance entities, or regulatory consulting organisations. The study was conducted independently within the BARU Pharmacoepidemiology Research Programme.

Data Availability Statement

Aggregate study-level RWEQI scores, regulatory acceptance outcomes, methodological feature data, and R analysis scripts are deposited at Zenodo:

<https://doi.org/10.5281/zenodo.19511917-data>.

Individual patient-level data from the Estonian EHIF and registry databases cannot be shared due to Estonian Personal Data Protection Act restrictions and data use agreement terms. Researchers may apply for data access directly to the EHIF Research Data Access Committee (research@haigekassa.ee).

Ethical Approval

All study analyses used de-identified population health data under ethics approval BARU-EC-2018-024 (Baltic AI Research University

Research Ethics Committee, Tallinn) and Estonian Health Insurance Fund Research Data Agreement EHIF-RDA-2018-04. Estonian Biobank data were accessed under institutional agreement EB-2019-12 with participants having provided broad consent for health research. All analyses comply with the Estonian Personal Data Protection Act and GDPR Article 9(2)(j) for scientific research purposes.

Appendix A

RWEQI Scoring Manual, TTE Protocol Template, and NCO Library Summary

This appendix provides the complete RWEQI sub-score computation rules, a target trial emulation protocol template for use in pharmacoepidemiology study planning, a curated negative control outcome library for Estonian EHR-claims studies, and worked RWEQI examples for high-quality and low-quality study designs.

A1. RWEQI Sub-Score Computation Rules

Data source quality: linked EHR-claims = 1.0; EHR only = 0.80; claims only = 0.70; registry = 0.64; patient-generated data = 0.54; all scores modifiable by missingness rate (deduct 0.10 per 10% missing on primary exposure or outcome).

Causal method: TTE = 0.884; MR IV (Biobank) = 0.884; geographic/time IV = 0.844; hdPS + E-value + NCO = 0.824; regression discontinuity = 0.804; D-in-D = 0.764; PS/IPTW only = 0.724; unadjusted = 0.284.

Confounding control completeness: +0.25 each for: active comparator design; NCO validation; E-value reported; covariate balance confirmed (e-SMD < 0.10); total 0-1.

Outcome validity: validated algorithm (PPV ≥ 0.85 in same or similar population) = 1.0; partially validated (PPV 0.70-0.85) = 0.70; expert review validation = 0.50; unvalidated = 0.40.

Regulatory compliance: FDA RWE Framework 2018 + EMA RWE Guidance 2021 checklist items completed / total items; supplemented by RECORD statement compliance (0-1).

A2. Negative Control Outcome Library: Top-10 Estonian Pharmacoepidemiology NCOs

1. Appendicitis (ICD-10 K37): biologically unrelated to most cardiovascular and metabolic drug exposures; incidence stable 0.12% annually; PPV ICD-10 code = 0.94 (Estonian validation).
2. Traumatic limb fracture (S42-S72): unrelated to most internal medicine drug exposures; detection complete in hospital claims; PPV 0.96.
3. Ingrown toenail (L60.0): unrelated to systemic medications; ascertained in GP + dermatology claims; PPV 0.88.
4. Inguinal hernia repair (K40-K41 + procedure code): unrelated to pharmacological exposures; complete capture in hospital register.
5. Cataract surgery (H26 + procedure): stable incidence; unrelated to most drug exposures except steroids; PPV 0.97.

6. Dental extraction (Z87.2 + procedure): unrelated to most systemic drug exposures; complete in claims.
7. Contact dermatitis NOS (L25.9): low confounder overlap with most systemic drug studies; PPV 0.82.
8. Benign skin lesion excision (L82 + procedure): unrelated to most exposures; complete surgical capture.
9. Wrist/hand sprain (S63): unrelated to most pharmacological exposures; complete emergency claims capture.
10. Acute sinusitis (J01): common; unrelated to most non-antibiotic drug exposures; PPV 0.84; useful for cardiovascular and metabolic drug NCO.