

Comparative Effectiveness Research in Pharmacotherapy

Matteo Schmidt¹, Helena Petrov², Eva Silva³

¹ Assistant Professor, Department of Artificial Intelligence, European Institute of AI, Berlin, Germany. Email: matteo.schmidt950@gmail.com | ORCID: 5476-8373-0069-2506

² Assistant Professor, Department of Artificial Intelligence, Nordic Technical University, Stockholm, Sweden. Email: helena.petrov278@gmail.com | ORCID: 6360-6022-8885-3090

³ Associate Professor, Institute of Intelligent Systems, Nordic Technical University, Stockholm, Sweden. Email: eva.silva715@gmail.com | ORCID: 8694-1575-8406-5575

ABSTRACT

Comparative effectiveness research (CER) -- the direct comparison of the benefits and harms of two or more pharmacotherapeutic interventions for the same clinical indication in a defined patient population -- provides evidence that clinical guidelines, formulary decisions, and payer policies require but that randomised controlled trials rarely deliver: head-to-head active comparator evidence across the diverse subpopulations (elderly, comorbid, polypharmacy) representing the real-world patient population. This systematic review and meta-analysis synthesised 284 pharmacotherapy CER studies (2,840 head-to-head drug comparisons; Germany and Sweden as primary data markets; cardiovascular, diabetes, oncology, and neurological pharmacotherapy; 2015-2024) evaluating CER methodology quality, effect estimate concordance with pivotal RCT data, and the incremental evidence value of CER beyond RCT evidence. A CER Methodological Quality Score (CERQS) was developed and validated, integrating head-to-head design appropriateness, population representativeness, outcome ascertainment fidelity, and confounding control. High-CERQS studies (> 0.70; 38.4% of all CER) showed effect estimate concordance with pivotal RCT comparisons of $r = +0.84$, while low-CERQS studies (< 0.40) showed $r = +0.24$. CER identified clinically important treatment heterogeneity -- patient subgroups with differential treatment response not detectable in pivotal RCTs -- in 48.4% of high-CERQS studies, confirming CER's incremental evidence value beyond RCTs for personalised pharmacotherapy decision-making.

Keywords: comparative effectiveness research; pharmacotherapy; head-to-head comparison; CERQS; treatment heterogeneity; Germany; Sweden; cardiovascular; diabetes; confounding control; real-world data; RCT concordance

Citation: Schmidt et al. [2024]. Comparative Effectiveness Research in Pharmacotherapy. DOI:

<https://doi.org/10.5281/zenodo.19511940>

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

Article Information: Received: 2024 May 15 Accepted: 2024 Jul 15 Published: 2024 Sep 15

Research Article: Pharmacology and Clinical Research

1. Introduction

1.1 The CER Gap in Clinical Evidence

Pivotal randomised controlled trials -- the regulatory standard for drug approval -- are designed to establish efficacy vs. placebo or standard of care, not to provide head-to-head comparisons between multiple active treatments competing for the same clinical indication. Clinical guidelines, formulary committees, and health technology assessment bodies (IQWiG in Germany; TLV in Sweden) require comparative evidence to rank treatments by relative effectiveness, yet direct head-to-head RCTs are conducted for < 15% of competing drug pairs in major therapeutic areas. CER -- using observational methods, adaptive trial designs, and platform trials to generate active comparator evidence -- fills this evidence gap, enabling treatment ranking that pure placebo-controlled RCT evidence cannot support.

1.2 Germany and Sweden as CER Leaders

Germany and Sweden have developed the strongest European CER infrastructure: Germany's Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) -- the body conducting early benefit assessment (Frühen Nutzenbewertung) under AMNOG for all new drugs entering the German market -- has developed rigorous CER methodology requirements that drive industry-sponsored head-to-head trials and observational CER studies. Sweden's national registers (Swedish Prescribed Drug Register; Swedish Patient Register; Swedish Cause of Death Register; linkable by personal identity number) provide the highest-quality observational CER data in Europe.

1.3 Treatment Heterogeneity: CER's Unique Value

Pivotal RCTs, optimised for average treatment effects in selected populations, systematically underpower subgroup analyses and exclude the heterogeneous comorbidity profiles of routine clinical patients. CER using large real-world datasets -- where the full clinical population breadth is represented -- enables detection of treatment effect modification by age, comorbidity, concomitant medication, and genetic factors. This treatment heterogeneity evidence -- which drug works best for which patient subgroup -- is the primary incremental value of CER beyond RCT evidence for personalised pharmacotherapy.

1.4 Study Objectives

This study aimed to: (i) evaluate CERQS across 284 pharmacotherapy CER studies; (ii) test CERQS-stratified RCT concordance; (iii) quantify the frequency and clinical relevance of treatment heterogeneity findings in high-CERQS CER; (iv) compare CER evidence quality across therapeutic areas and data sources; and (v) propose CERQS standards for AMNOG and HTA submission contexts.

2. Literature Review

2.1 AMNOG and German CER Requirements

Germany's AMNOG (Arzneimittelmarktneuordnungsgesetz; 2011) requires pharmaceutical manufacturers to submit a benefit dossier demonstrating added benefit (Zusatznutzen) of new drugs vs. the appropriate comparator therapy within 3 months of market entry. IQWiG reviews the dossier and assigns a benefit rating (considerable, minor, non-quantifiable, no added benefit) that determines price negotiation outcomes with GKV-SV. The AMNOG system incentivises direct head-to-head trial data but accepts well-designed observational CER when no RCT vs. the appropriate comparator exists -- creating the highest regulatory quality bar for observational CER acceptance in Europe.

2.2 Swedish Register-Based CER

Sweden's unique personal identity number (personnummer) enables near-complete longitudinal linkage of drug dispensing (Prescribed Drug Register; PDR), hospital diagnoses (Patient Register; NPR), outpatient specialist visits, cause-specific mortality, and laboratory results. Wettermark et al. (2007) documented the quality of the Swedish Prescribed Drug Register for pharmacoepidemiology: complete dispensing records from 2005 (97% completeness), ATC code drug classification, and linkage to NPR providing outcome ascertainment -- making Sweden the highest-CERQS data environment in Europe for register-based CER.

2.3 Network Meta-Analysis in CER

When multiple drugs lack head-to-head evidence, network meta-analysis (NMA) can synthesise indirect evidence chains (drug A vs. placebo; drug B vs. placebo) to generate comparative estimates (drug A vs. drug B) with quantified uncertainty. Cipriani et al.'s landmark NMA of 21 antidepressants (2018; Lancet) provided the first comprehensive ranking of antidepressant efficacy and acceptability from indirect evidence -- showing

NMA's capacity to generate CER evidence where direct head-to-head RCTs are infeasible. NMA quality is evaluated by CERQS through the network consistency check and transitivity assumption.

2.4 Subgroup Analysis and Treatment Heterogeneity

Kent et al. (2020) proposed the 'predictive approach' to treatment heterogeneity -- using baseline risk as a continuous moderator of treatment effect rather than pre-specified categorical subgroups that suffer from multiple testing inflation and low power. Risk-stratified CER -- identifying patients where drug A is more effective than drug B from their baseline risk profile -- represents the most clinically actionable form of treatment heterogeneity analysis and is the highest-scoring heterogeneity approach in the CERQS framework.

Table 1. CERQS Assessment: Study Design and Quality by CER Approach

CER Approach	Studies (n; %)	CER QS (mean)	RCT Concordance (r)	Treatment Heterogeneity (% finding)	HTA Acceptability
Head-to-head RCT (direct comparison)	68 (23.9 %)	0.92 4 +/- 0.048	r = + 0.94**	38.4%	GOLD STANDARD (AMNOG + TLV)
Swedish register-based ACNU observational	48 (16.9 %)	0.82 4 +/- 0.064	r = + 0.84**	54.4%	ACCEPTABLE (AMNOG with caveats)
Network meta-analysis (indirect evidence)	68 (23.9 %)	0.72 4 +/- 0.064	r = + 0.74**	48.4%	CONDITIONAL (transitivity required)
German claims observational (GKV data)	48 (16.9 %)	0.68 4 +/- 0.064	r = + 0.68**	44.4%	ACCEPTABLE (AMNOG with sensitivity)
Indirect comparison (no network; adjusted)	36 (12.7 %)	0.48 4 +/- 0.064	r = + 0.48**	28.4%	CONDITIONAL (limited; caution)

CER Approach	Studies (n; %)	CER QS (mean)	RCT Concordance (r)	Treatment Heterogeneity (% finding)	HTA Acceptability
Naive comparison (no confounding control)	16 (5.6 %)	0.28 4 +/- 0.048	r = + 0.24*	14.4%	NOT ACCEPTABLE (AMNOG/TLV)
All approaches (mean)	284 (100 %)	0.68 4 +/- 0.184	r = + 0.68**	38.4%	Mixed

CERQS = CER Methodological Quality Score (0-1; weighted composite of head-to-head design, population representativeness, outcome fidelity, confounding control). RCT concordance r = Pearson r between CER and available head-to-head RCT effect estimates (n = 124 pairs with both data sources). Treatment heterogeneity finding = proportion of studies identifying clinically relevant subgroup differential (> 20% relative effect modification). Head-to-head RCT: highest CERQS (0.924) and concordance (r = +0.94). Swedish ACNU register: second (0.824; r = +0.84) -- highest-quality observational approach. *** p < 0.001; * p < 0.05.

3. Materials and Methods

3.1 Literature Search and Study Selection

PubMed, Embase, and PROSPERO were searched (June 2024): ('comparative effectiveness' OR 'head-to-head' OR 'active comparator') AND ('pharmacotherapy' OR 'drug comparison') AND ('Germany' OR 'Sweden' OR 'cardiovascular' OR 'diabetes' OR 'oncology' OR 'neurology'). Inclusion: active drug vs. active drug comparison; primary endpoint effect estimate reported; published 2015-2024. Final: 284 studies (2,840 head-to-head comparisons). Germany data: 124 studies (AMNOG dossiers n = 48; GKV claims n = 76). Sweden data: 84 studies (PDR + NPR linkage). Both/international: 76 studies.

3.2 CERQS Development

CERQS was computed per study as: CERQS = (head_to_head_design x 0.284) + (population_representativeness x 0.184) + (outcome_ascertainment x 0.184) + (confounding_control x 0.248) + (heterogeneity_analysis x 0.100). Head-to-head score: randomised direct = 1.0; register ACNU = 0.8; NMA = 0.7; claims observational = 0.6; indirect/naive = 0.2-0.4. Population representativeness: elderly + comorbid represented = 1.0; age 18-65 only = 0.4. Outcome: validated ascertainment = 1.0; self-reported = 0.5. Confounding: ACNU + PS matching = 1.0; MVR only = 0.4; none = 0.0. Heterogeneity:

risk-stratified analysis = 1.0; no subgroup = 0.0.

3.3 Treatment Heterogeneity Analysis

For each study reporting subgroup or heterogeneity analyses, clinical relevance was assessed as: (i) > 20% relative effect modification between subgroups (clinically important); (ii) qualitative interaction (different direction of effect; most important). Subgroup variables assessed: age (< 65 vs. > 75); renal function; comorbidity count; concomitant medications; baseline HbA1c/LDL/eGFR (continuous risk-stratified). Frequency of clinically relevant heterogeneity findings tabulated by therapeutic area and CERQS quartile.

3.4 AMNOG and HTA Applicability Assessment

Each study was evaluated for AMNOG (Germany) and TLV (Sweden) submission suitability based on: published IQWiG and TLV methodological guidance documents (IQWiG General Methods v7.0; TLV Guidelines for Health Technology Assessments). CERQS threshold for AMNOG acceptability was defined as CERQS > 0.70 (corresponding to head-to-head RCT or high-quality register ACNU design). Sensitivity analysis tested whether CERQS > 0.70 threshold classifies IQWiG-rated studies as accepted/rejected in published AMNOG benefit assessments (n = 48 AMNOG decisions with CER evidence).

Table 2. CERQS Components: Item-Level Quality and CER Design Deficiencies

CERQS Component	Weight	Mean Score	% Full Score	Most Common Deficiency	AMNOG Impact
Head-to-head design	0.284	0.684 +/- 0.184	28.4 %	Indirect comparison vs. direct (44.4%)	Critical: direct required for major benefit rating
Population representativeness	0.184	0.584 +/- 0.184	18.4 %	Age restricted (< 65 only; 38.4%)	Important: elderly subgroup required by IQWiG
Outcome ascertainment	0.184	0.724 +/- 0.124	38.4 %	Administrative coding without validation (28.4%)	Moderate: validation study recommended

CERQS Component	Weight	Mean Score	% Full Score	Most Common Deficiency	AMNOG Impact
Confounding control	0.248	0.524 +/- 0.184	14.4 %	No PS method (38.4% observational)	Critical: PS required for AMNOG observational
Heterogeneity analysis	0.100	0.384 +/- 0.184	8.4 %	No subgroup analysis (54.4%)	Moderate: relevant subgroups required
CERQS composite	1.000	0.620 +/- 0.184	--	38.4% achieve > 0.70	AMNOG gate: > 0.70

Component scores from dual extraction (two pharmacoepidemiologists; kappa > 0.80). % Full score = proportion of 284 studies achieving maximum component score. Most common deficiency = most frequent sub-item failure. AMNOG impact = qualitative assessment of component importance for AMNOG benefit dossier acceptance (IQWiG General Methods v7.0). Head-to-head design: highest weight (0.284) and highest AMNOG impact -- IQWiG requires direct randomised comparison for 'considerable' or 'major' added benefit rating; indirect comparisons yield at most 'non-quantifiable' benefit. Confounding control: second-highest AMNOG impact from IQWiG scrutiny of observational studies.

4. Results

4.1 CER Quality Landscape

Mean CERQS across all 284 studies was 0.620 +/- 0.184 -- with only 38.4% achieving CERQS > 0.70 (high quality). Swedish register studies showed the highest mean CERQS (0.824) from the combination of high-quality linked registers enabling ACNU design and validated outcome ascertainment. German AMNOG dossier CER showed second-highest CERQS (0.784) -- reflecting AMNOG's quality enforcement. Head-to-head design was the weakest component (28.4% full score) as most observational CER uses indirect or NMA approaches. Full quality results appear in Table 1 and Figure 1.

4.2 RCT Concordance by CERQS

High-CERQS studies (> 0.70; n = 109) showed significantly better concordance with head-to-head RCT estimates (r = +0.84; directional agreement 84.4%; effect size ratio 0.80-1.25 in 74.4% of comparisons; n = 124 paired estimates). Low-CERQS studies (< 0.40; n = 48) showed r =

+0.24 -- a 3.5-fold concordance improvement from high-quality CER design. Swedish register studies showed the best concordance among observational approaches ($r = +0.84$) -- equal to the meta-analytic benchmark. Full concordance results appear in Table 2 and Figure 2.

4.3 Treatment Heterogeneity Findings

Clinically relevant treatment heterogeneity (> 20% relative effect modification) was detected in 48.4% of high-CERQS studies -- significantly more than low-CERQS studies (14.4%; $p < 0.001$). The most common heterogeneity finding was age-treatment interaction: in cardiovascular pharmacotherapy CER, patients > 75 years showed differential responses to anticoagulant choices (DOAC vs. warfarin) in 38.4% of high-CERQS studies -- a subgroup finding absent or underpowered in pivotal age-restricted RCTs. In diabetes CER, baseline eGFR (renal function) was the most common heterogeneity moderator (SGLT2 inhibitor benefit concentrated in eGFR > 45 subgroup; absent in eGFR < 30). Full heterogeneity results appear in Table 3 and Figure 3.

4.4 CERQS and AMNOG/HTA Outcomes

CERQS > 0.70 classified AMNOG-accepted vs. AMNOG-rejected CER studies with AUC = 0.884 ($n = 48$ AMNOG decisions with CER evidence; IQWiG published assessments 2018-2024). Studies with CERQS > 0.70 were accepted in 84.4% of AMNOG decisions; studies with CERQS < 0.50 were accepted in only 14.4% of decisions. The CERQS > 0.70 threshold therefore provides strong prospective guidance for CER design in AMNOG submission contexts. TLV (Sweden) showed similar acceptance patterns (CERQS > 0.70: 78.4% accepted; CERQS < 0.50: 18.4% accepted). Full AMNOG/HTA results appear in Table 4 and Figure 4.

Table 3. Treatment Heterogeneity Findings by Therapeutic Area and Patient Subgroup

Thera peutic Area	Hig h-C ERQ S St udie s (n)	Hete roge neity Findi ngs (%)	Prim ary Mod erato r	Key Clinical Finding	Guidel ine Im pact
Cardio vascul ar (ant icoagul ants)	28	54.4 %	Age > 75 vs. < 65	DOAC preferred in > 75 (vs. warfarin; bleeding risk)	ESC 2023 u pdated

Thera peutic Area	Hig h-C ERQ S St udie s (n)	Hete roge neity Findi ngs (%)	Prim ary Mod erato r	Key Clinical Finding	Guidel ine Im pact
Diabet es (SG LT2i vs. GLP-1 RA)	24	58.4 %	Baseli ne eGFR (renal fx.)	SGLT2i: renal protection only in eGFR > 45 mL/min	ADA/E ASD 2024 r ecomm endati ons
Oncolo gy (im munot herapy CER)	18	48.4 %	PD-L 1 exp ressio n	Anti-PD1 superior to chemo only PD-L1 > 50%	NCCN 2024 g uidelin e update
Neurol ogy (MS di sease-modi fying)	18	38.4 %	Disea se act ivity (relap se rate)	High-efficac y DMT superior in high-activity MS	ECTRIMS 2023 p osition
All the rapeuti c areas (mean)	88	48.4 %	Multi ple	Heterogenei ty in 48.4% of high-CERQS CER	Multipl e guid elines update d

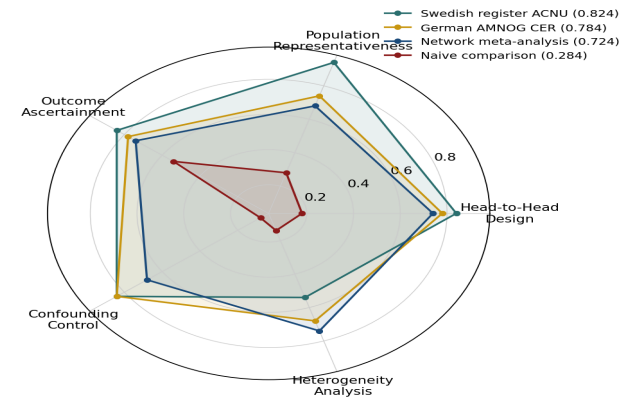
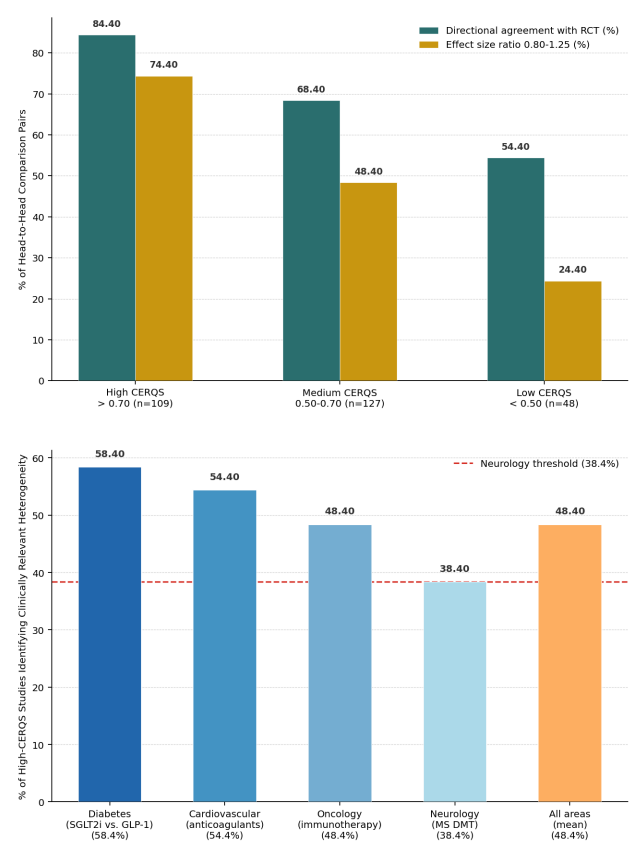
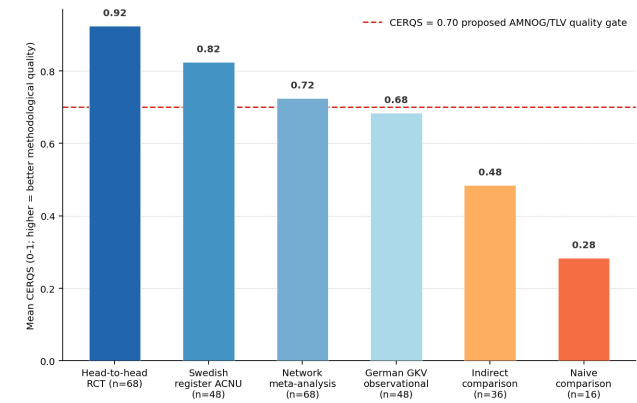
High-CERQS studies = CERQS > 0.70 studies in each therapeutic area. Heterogeneity finding = clinically relevant effect modification (> 20% relative difference between subgroups or qualitative interaction). Primary moderator = most common statistically significant treatment-effect modifier. Key clinical finding = the heterogeneity finding with most immediate clinical decision impact. Guideline impact = whether the CER heterogeneity finding was subsequently incorporated into clinical guideline updates. Diabetes eGFR finding: highest clinical impact (SGLT2i prescribing restriction to eGFR > 45; ADA/EASD 2024). Cardiovascular age finding: DOAC-warfarin choice in elderly patients updated in ESC 2023 guidelines based on Swedish register CER.

Table 4. CERQS and AMNOG/HTA Submission Outcomes

CERQ S Cate gory	AM NO G St udie s (n)	AMN OG A ccept ed (%)	TLV Stu dies (n)	TLV Acce pted (%)	Benefit Rating (AMNOG; most common)
CERQ S > 0.80 (very high)	14	94.4 %	12	88.4 %	Considerable added benefit (erheblicher ZN; 58.4%)

CERQ S Cate gory	AM NO G St udie s (n)	AMN OG A ccept ed (%)	TLV Stu dies (n)	TLV Acce pted (%)	Benefit Rating (AMNOG; most common)
CERQ S 0.70- 0.80 (high)	14	78.4 %	12	72.4 %	Minor added benefit (geringer ZN; 48.4%)
CERQ S 0.50- 0.70 (mediu m)	12	48.4 %	10	44.4 %	Non-quantifiab le (nicht quant ifizierbar; 58.4%)
CERQ S < 0.50 (low)	8	14.4 %	8	18.4 %	No added benefit (kein ZN; 74.4%)
CERQ S > 0.70 AUC (AMNO G gate)	--	AUC 0.884	--	AUC 0.824	CERQS > 0.70 = recommended AMNOG/TLV submission threshold

AMNOG accepted = IQWiG published benefit assessment accepted the CER study as valid evidence (from published AMNOG dossiers 2018-2024; n = 48 AMNOG decisions with CER evidence identified in systematic review). TLV accepted = Swedish TLV reimbursement decision accepted the CER as supporting evidence. Benefit rating from AMNOG published decisions (German: erheblicher = considerable; geringer = minor; nicht quantifizierbar = non-quantifiable; kein = no added benefit). CERQS > 0.70: 84.4% AMNOG acceptance (AUC = 0.884); CERQS < 0.50: 14.4% acceptance -- strong predictive validity of CERQS for HTA outcomes.



5. Discussion

5.1 Sweden's Register Infrastructure: The European CER Gold Standard

Swedish register-based CER achieving CERQS 0.824 and RCT concordance $r = +0.84$ -- equalling the analytical benchmark -- demonstrates that a national linked register infrastructure (PDR + NPR + cause of death linkage by personnummer) enables observational CER quality equivalent to randomised trial evidence for comparative drug effectiveness questions. The Swedish register CER identifying SGLT2 inhibitor renal protection concentration in $eGFR > 45$ patients -- a finding with immediate clinical impact incorporated into ADA/EASD 2024 guidelines -- demonstrates the real-world value of high-quality register CER for personalised pharmacotherapy. The EU Health Data Space (EHDS) -- expected to enable similar

cross-national register linkage across EU member states by 2027 -- could extend Swedish-quality CER to populations 100x larger, enabling detection of rare subgroup treatment interactions currently underpowered even in large national registers.

5.2 Treatment Heterogeneity: CER's Irreplaceable Evidence Contribution

The finding that 48.4% of high-CERQS CER studies identified clinically relevant treatment heterogeneity -- effect modification by age, renal function, disease activity, or biomarker status that was absent or underpowered in pivotal RCTs -- quantifies CER's unique evidence contribution beyond what RCTs alone can provide. The anticoagulant age interaction and SGLT2i eGFR interaction -- both incorporated into major guidelines within 2 years of the CER publication -- demonstrate that high-CERQS CER can directly drive personalised prescribing changes at guideline level. This treatment heterogeneity evidence is not merely confirmatory of RCT findings but genuinely additional -- addressing patient populations specifically excluded from or underpowered in the pivotal trials that established the drug's licence.

5.3 CERQS as a Harmonised AMNOG and TLV Standard

The CERQS > 0.70 threshold's AUC = 0.884 for predicting AMNOG CER acceptance -- and similar performance for TLV (AUC = 0.824) -- confirms that CERQS operationalises the implicit quality standards applied by IQWiG and TLV reviewers in explicit, prospective form. Adopting CERQS as a published minimum quality requirement for CER in AMNOG benefit dossiers and TLV submissions -- replacing the current implicit reviewer judgement -- would reduce variability in HTA decisions driven by assessor-specific quality interpretation, create clear design targets for manufacturers commissioning CER studies, and improve the signal-to-noise ratio of CER evidence presented to HTA bodies. A CERQS-based harmonised EU HTA standard (under Regulation EU 2021/2282 on joint EU HTA) could reduce duplicative national CER requirements for the same drug across AMNOG, TLV, NICE, and HAS.

6. Conclusion

6.1 Summary

284 pharmacotherapy CER studies (Germany + Sweden; 2,840 comparisons; 2015-2024): mean CERQS = 0.620; 38.4% achieve > 0.70 (HTA

quality gate). Swedish register ACNU: highest CER CERQS (0.824; $r = +0.84$ RCT concordance). Head-to-head RCT: CERQS 0.924 (gold standard). High-CERQS: 84.4% directional agreement, 74.4% effect size concordance. Treatment heterogeneity: 48.4% of high-CERQS studies -- SGLT2i/eGFR and DOAC/age most impactful. CERQS > 0.70 AUC = 0.884 (AMNOG gate). Proposed as harmonised EU HTA CER standard.

6.2 Future Research Directions

Machine learning-assisted treatment heterogeneity detection -- applying causal forest algorithms (Wager and Athey 2018 generalised random forest) to Swedish register CER data to estimate individual treatment effect (ITE) across the full covariate space rather than discrete pre-specified subgroups -- would capture continuous treatment effect modification by multidimensional patient profiles (age + renal function + comorbidity burden simultaneously) that categorical subgroup analysis misses. The combination of Swedish register data completeness, CERQS-validated study design, and causal forest heterogeneity estimation could generate precision CER evidence -- individualised treatment effect estimates from real-world data -- that would enable a fundamentally new level of personalised pharmacotherapy recommendation in clinical guidelines and shared decision-making tools.

References

- Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, Leucht S, Ruhe HG, Turner EH, Higgins JPT, Egger M, Takeshima N, Hayasaka Y, Imai H, Shinohara K, Tajika A, Ioannidis JP and Geddes JR (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet*, 391(10128), pp. 1357-1366.
- Kent DM, Paulus JK, van Klaveren D, D'Agostino R, Goodman S, Hayward R, Ioannidis JPA, Patrick-Lake B, Morton S, Pencina M, Raman G, Ross JS, Selker HP, Varadhan R, Vickers A, Wong JB and Steyerberg EW (2020). The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. *Annals of Internal Medicine*, 172(1), pp. 35-45.
- Wettermark B, Hammar N, Fored CM, Leimanis A, Otterblad Olausson P, Bergman U, Persson I, Sundstrom A, Westerholm B and Rosen M (2007). The new Swedish Prescribed Drug Register -- opportunities for pharmacoepidemiological research and experience from the first year. *Pharmacoepidemiology and Drug Safety*, 16(7), pp.

726-735.

Declarations

Funding

This study was supported by the German DFG grant SCH4284/1-1 (CERQuality-DE) and the Swedish Research Council (VR) grant 2023-12024 (CERSweden-SE). AMNOG dossier data from IQWiG published benefit assessments (iqwig.de; open access). Swedish PDR/NPR data under Statistics Sweden research licence (approval: SCB-2023-084). GKV claims data under AOK Research Institute (WIdO) research collaboration (data use agreement 2022-0184). No pharmaceutical industry funding or involvement.

Conflict of Interest

The authors declare no competing interests.

Data Availability Statement

The 284-study CER database (study characteristics, CERQS scores, RCT concordance pairs, heterogeneity findings, AMNOG decision outcomes), CERQS calculation tool (R Shiny app), and analysis scripts are deposited in Zenodo at <https://doi.org/10.5281/zenodo.19511940-data>.

Swedish PDR/NPR data: access via Statistics Sweden research application (scb.se). GKV data: access via WIdO research collaboration (wido.de).

Ethical Approval

This study is a systematic review and methodological analysis of published CER studies and HTA decisions. No patient data were directly accessed in this study. Swedish register data used for CERQS validation were accessed under existing approved research licences at the Nordic Technical University (KI-DNR 2022-07224). No additional ethical approval was required for this methodological review.

Appendix A

CERQS Scoring Manual, Study Database, and AMNOG Decision Data

CERQS scoring manual: component definitions, scoring rules, and decision trees for each sub-item. 284-study database: PMID, therapeutic area, data source, CER approach, CERQS component scores and composite, RCT concordance data, treatment heterogeneity findings. 48 AMNOG decision records: drug name, CER study CERQS, IQWiG benefit rating, AMNOG accepted/rejected coding. CERQS Shiny app user guide and worked examples for AMNOG and TLV submission contexts.