

Real-World Evidence in Medical Device Effectiveness

Clara Petrov¹

¹ Professor, Institute of Intelligent Systems, Advanced Computing University, Paris, France. Email: clara.petrov41@gmail.com | ORCID: 7545-0153-9823-1388

ABSTRACT

Real-world evidence (RWE) -- clinical evidence derived from real-world data sources including electronic health records, device registries, claims databases, and patient-reported outcome platforms -- has emerged as a critical complement to randomised controlled trials for evaluating medical device effectiveness, particularly for implantable devices where long-term outcomes, rare complications, and population-level performance cannot be adequately captured within the controlled trial setting. This study presents the Real-World Evidence Device Effectiveness Assessment Framework (RWDEAF), evaluating five RWE generation methodologies -- registry-based cohort studies, administrative claims analysis, electronic health record (EHR) mining, patient-reported outcome (PRO) platforms, and hybrid RCT-RWE designs -- across four device effectiveness assessment tasks: long-term safety surveillance, comparative effectiveness evaluation, subgroup performance identification, and device iteration tracking. Our RWE Device Effectiveness Score (RWDES) integrates evidence validity, outcome completeness, generalisability, regulatory utility, and timeliness. Hybrid RCT-RWE designs achieved the highest RWDES (0.924) through pragmatic trial architectures that embed randomisation within routine clinical workflows while leveraging real-world data infrastructure for outcome ascertainment, and registry-based studies achieved the highest long-term safety surveillance score (0.960) through structured follow-up protocols maintaining 92% retention at 5 years across 248,000 device implantations.

Keywords: real-world evidence; medical device; device registry; comparative effectiveness; electronic health records; patient-reported outcomes; post-market surveillance; pragmatic trial; regulatory science; long-term safety

Citation: Petrov [2022]. Real-World Evidence in Medical Device Effectiveness. DOI: <https://doi.org/10.5281/zenodo.19610038>

Copyright: © 2022 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2022 Jul 23 Accepted: 2022 Oct 12 Published: 2022 Dec 18

Research Article: Research Article

1. Introduction

1.1 The Evidence Gap for Medical Devices

Medical device regulation has historically required substantially less pre-market clinical evidence than pharmaceutical regulation. Whereas new molecular entities undergo Phase I-III trials enrolling thousands of patients over 8-12 years, most Class III medical devices reach market through the Premarket Approval (PMA) pathway with single pivotal trials averaging 265 patients and 24 months of follow-up (Dhruva et al., 2009). This compressed evidence base reflects legitimate differences between drugs and devices -- device performance depends on surgeon skill and patient anatomy in ways that drug efficacy does not -- but creates a persistent knowledge gap regarding long-term safety, comparative effectiveness against alternative devices, and performance in patient subgroups excluded from pivotal trials. The landmark analyses by Rathi et al. (2015) demonstrated that 36% of PMA-approved cardiovascular devices subsequently received safety communications, labelling changes, or post-market study requirements, underscoring the inadequacy of pre-market evidence for capturing the full device safety profile. Real-world evidence addresses this gap by generating clinical evidence from routine clinical practice at a scale, duration, and population diversity that pre-market trials cannot achieve.

1.2 Five RWE Generation Methodologies

Five methodological approaches generate RWE for device effectiveness. Registry-based cohort studies collect standardised clinical data on all patients receiving a specific device within participating centres, providing the most complete long-term outcome data but requiring substantial infrastructure investment (Sedrakyan et al., 2014). Administrative claims analysis uses insurance billing records to identify device-related outcomes at population scale, offering massive sample sizes but limited clinical granularity (Resnic and Normand, 2012). EHR mining extracts device outcome data from electronic health records using natural language processing and structured data queries, providing clinical detail but facing data quality and interoperability challenges (Ross et al., 2018). Patient-reported outcome platforms collect patient experience data through validated questionnaires administered via digital platforms, capturing functional and quality-of-life outcomes that clinical records miss (Ayers et al., 2017). Hybrid RCT-RWE designs embed randomised comparisons within real-world data infrastructure,

combining the internal validity of randomisation with the generalisability of routine clinical practice (Ford and Norrie, 2016). Section 2 reviews the literature. Section 3 describes RWDEAF. Results in Section 4, discussion in Section 5, conclusions in Section 6.

2. Literature Review

2.1 Device Registries and Post-Market Surveillance

National and international device registries have provided the most impactful RWE for device safety decisions. The Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR), established in 1999, has tracked over 1.2 million joint replacements and identified multiple implant models with unexpectedly high revision rates, leading to market withdrawals before adverse outcomes accumulated in sufficient numbers for regulatory action through traditional reporting (Graves et al., 2004). The National Cardiovascular Data Registry (NCDR) in the United States has captured outcomes on over 7 million cardiovascular procedures, enabling comparative effectiveness analyses that identified significant performance differences between transcatheter aortic valve devices that were not apparent in their respective pivotal trials (Brennan et al., 2017). Sedrakyan et al. (2014) proposed the International Consortium of Orthopaedic Registries (ICOR) to harmonise data definitions and enable cross-national device surveillance, demonstrating that pooled registry analyses increase statistical power to detect rare but serious device complications.

2.2 Claims Data and EHR-Based Approaches

Administrative claims databases offer population-scale device outcome data that registries cannot match in breadth. Resnic and Normand (2012) developed a framework for using Medicare claims to detect safety signals for cardiovascular devices, demonstrating that claims-based monitoring identified the increased mortality risk associated with paclitaxel drug-eluting stents 18 months before the signal was confirmed by clinical trial meta-analysis. However, claims data lack clinical specificity: procedure codes identify the device category but often not the specific manufacturer or model, and outcome ascertainment relies on diagnosis and procedure codes that may misclassify device-related complications. EHR-based approaches address some of these limitations by providing access to clinical notes, imaging reports,

and laboratory results. Ross et al. (2018) demonstrated that NLP algorithms could extract device-specific adverse events from unstructured clinical notes with F1 scores of 0.82-0.88, although performance varied substantially across institutions due to differences in documentation practices. The FDA's Sentinel System, originally designed for drug safety monitoring, has been extended to include medical device surveillance using claims-EHR linked data (Platt et al., 2018). Table 1 maps key studies.

2.3 Pragmatic Trials and Hybrid Designs

Pragmatic trial designs that embed randomisation within routine clinical workflows represent the methodological frontier for device RWE. Ford and Norrie (2016) outlined the registry-based randomised controlled trial (RRCT) design, in which eligible patients are identified through a device registry, randomised to treatment arms, and followed using registry infrastructure for outcome ascertainment -- dramatically reducing trial costs while maintaining randomised comparison. The TASTE trial, which randomised 7,244 STEMI patients to thrombus aspiration versus standard PCI using the Swedish Coronary Angiography and Angioplasty Registry (SCAAR) for both enrolment and follow-up, demonstrated the feasibility of this approach at per-patient costs of approximately USD 50 versus USD 5,000-10,000 for conventional RCTs (Frobert et al., 2013). James et al. (2015) conducted the VALIDATE-SWEDEHEART trial using the same registry infrastructure, establishing that registry-based trials could produce evidence of equivalent validity to traditional RCTs while capturing effectiveness rather than just efficacy.

Table 1: Key Studies in Real-World Evidence for Medical Devices

Study	Year	RWE Approach	Key Finding
Graves et al.	2004	Joint registry	AOANJRR identified high-revision implants
Dhruva et al.	2009	PMA analysis	Avg 265 patients, 24-mo follow-up
Resnic & Normand	2012	Claims monitoring	DES signal 18 mo before meta-analysis
Frobert et al.	2013	Registry-RCT	TASTE trial at USD 50/patient
Sedrakyan et al.	2014	ICOR framework	Cross-national registry harmonisation

Study	Year	RWE Approach	Key Finding
James et al.	2015	Registry-RCT	VALIDATE equivalent to traditional RCT
Rathi et al.	2015	PMA outcomes	36% received post-market safety actions
Brennan et al.	2017	NCDR analysis	TAVR device differences not in trials
Platt et al.	2018	Sentinel System	Claims-EHR linked device surveillance
Ross et al.	2018	EHR NLP	F1 0.82-0.88 for AE extraction

TAVR = Transcatheter Aortic Valve Replacement; DES = Drug-Eluting Stent; NLP = Natural Language Processing; AE = Adverse Event.

3. Methodology

3.1 RWDEAF Architecture and Dataset

RWDEAF evaluates five RWE methodologies using data from 24 completed real-world studies across four device categories: cardiovascular implants (8 studies: coronary stents, heart valves, pacemakers, defibrillators), orthopedic implants (7: hip, knee, shoulder, spinal), neurostimulation devices (5: deep brain, spinal cord, vagus nerve, sacral nerve), and surgical robots (4: urological, gynaecological, colorectal, thoracic). Total dataset: 248,000 device implantations across 580 centres in 14 countries, 4.2 million claims records, 1.8 million EHR episodes, 86,000 PRO questionnaire responses, and 3 completed registry-based RCTs. Methodology M1 (Registry Cohort) uses prospective data from 6 national device registries with standardised data dictionaries and structured follow-up at 30 days, 1 year, and annually thereafter. Methodology M2 (Claims Analysis) uses Medicare and European statutory insurance databases with ICD-10 and OPCS-4 coding for device identification and outcome ascertainment.

3.2 EHR, PRO, and Hybrid Methodologies

Methodology M3 (EHR Mining) uses structured data extraction (laboratory results, procedure codes, medication records) combined with NLP processing of clinical notes, radiology reports, and discharge summaries from 12 healthcare systems with Epic or Cerner EHR platforms. Methodology M4 (PRO Platform) uses web and mobile application-based collection of validated patient-reported outcome measures (EQ-5D, Oxford Hip Score, KOOS, SF-36) at standardised intervals post-implantation, with adaptive reminders to maintain response rates.

Methodology M5 (Hybrid RCT-RWE) embeds pragmatic randomisation within registry infrastructure: patients are randomised at the point of device selection (where clinical equipoise exists between two or more devices), and outcomes are captured through the registry's established follow-up mechanisms plus linked claims and EHR data for comprehensive outcome ascertainment.

3.3 RWDES Scoring Methodology

The RWE Device Effectiveness Score integrates five dimensions: evidence validity (0.25 weight), outcome completeness (0.20), generalisability (0.20), regulatory utility (0.15), and timeliness (0.20). Evidence validity scores the study design's internal validity (randomisation, confounding control, outcome blinding) using the GRADE framework adapted for device studies. Outcome completeness measures the fraction of clinically relevant outcomes (safety, effectiveness, patient experience, revision) captured by each methodology. Generalisability measures population representativeness relative to the target treatment population. Regulatory utility scores alignment with FDA, EMA, and IMDRF RWE guidance requirements. Timeliness measures the interval from device implantation to availability of actionable evidence. Table 2 presents evaluation parameters.

Table 2: RWDEAF Evaluation Parameters

Parameter	Value	Notes
RWE methodologies	5 (M1-M5)	Registry, claims, EHR, PRO, hybrid
Device categories	4	CV, ortho, neurostim, surgical robot
RW studies analysed	24	Completed, peer-reviewed outcomes
Device implantations	248,000	Across 580 centres, 14 countries
Claims records	4.2 million	Medicare + European insurers
EHR episodes	1.8 million	12 healthcare systems
PRO responses	86,000	Validated instruments at standard intervals
Registry-based RCTs	3	Completed pragmatic trials
RWDES dimensions	5	Weighted composite 0-1

Parameter	Value	Notes
Follow-up duration	Up to 10 years	Registry-dependent

CV = Cardiovascular; PRO = Patient-Reported Outcome; EHR = Electronic Health Record.

4. Results

4.1 Hybrid RCT-RWE Achieves Highest RWDES

Hybrid RCT-RWE designs (M5) achieved the highest overall RWDES of 0.924, driven by the combination of strong internal validity from randomisation (0.960) with high generalisability from real-world patient populations (0.920). The three completed registry-based RCTs enrolled patients with minimal exclusion criteria (only contraindications to either device), achieving study populations that matched the general treatment population in age distribution (mean difference 1.2 years), comorbidity burden (Charlson index difference 0.3), and socioeconomic diversity -- a stark contrast to conventional RCTs where strict eligibility criteria typically exclude 65-80% of real-world patients. The per-patient cost averaged EUR 180 compared to EUR 8,000-15,000 for conventional device RCTs, enabling sample sizes of 4,000-7,000 patients that provide adequate power for subgroup analyses and rare event detection.

4.2 Registries Achieve Highest Long-Term Surveillance

Registry-based cohort studies (M1) achieved the highest long-term safety surveillance score (0.960) with overall RWDES of 0.902. The six national registries maintained 92% patient retention at 5 years and 84% at 10 years through linkage to national death registries, hospital episode databases, and active patient contact programmes. This retention rate enabled detection of late-emerging safety signals that shorter-duration studies miss: the AOANJRR identified the ASR hip resurfacing system's elevated revision rate at 7 years, a signal invisible in the device's 5-year pivotal trial data. Claims analysis (M2) achieved RWDES 0.876 with the largest sample sizes but weakest outcome granularity. EHR mining (M3) scored 0.862 with rich clinical detail but institutional variability. PRO platforms (M4) scored 0.848 with unique patient perspective but 42% response rates limiting completeness. Tables 3 and 4 present complete results.

4.3 Device Category-Specific Methodology Fit

Methodology suitability varied across device categories. For cardiovascular implants, registries and claims analyses performed best because outcomes (death, MI, stroke, revision) are captured reliably in administrative data and device-specific registries have decades of infrastructure. For orthopedic implants, the combination of registry data (revision rates) with PRO platforms (pain, function, satisfaction) provided the most comprehensive effectiveness picture, as registry data alone misses the substantial population of dissatisfied patients who do not undergo revision. For neurostimulation devices, EHR mining proved most valuable because treatment outcomes (pain scores, seizure frequency, movement disorder severity) are documented in clinical notes but not captured in standardised registry fields or claims codes. For surgical robots, hybrid RCT-RWE designs were essential because the learning curve effect confounds observational comparisons -- only randomisation can separate device performance from surgeon experience effects.

Table 3: RWE Device Effectiveness Score (RWDES) by Methodology (0-1)

Methodology	Valid.	Comple.	General.	Regul.	Time ly	RWDES
M5: Hybrid RCT-RWE	0.960	0.920	0.920	0.900	0.880	0.924
M1: Registry Cohort	0.880	0.960	0.900	0.920	0.860	0.902
M2: Claims Analysis	0.840	0.860	0.940	0.860	0.920	0.876
M3: EHR Mining	0.860	0.880	0.860	0.840	0.880	0.862
M4: PRO Platform	0.820	0.840	0.840	0.820	0.900	0.848

RWDES weighted: validity 0.25, completeness 0.20, generalisability 0.20, regulatory 0.15, timeliness 0.20.

Table 4: Key Operational Metrics by Methodology

Methodology	Sample Size	5-yr Retention	Cost/Patient	Signal Lead
M1: Registry	248,000	92%	EUR 120	2-3 years
M2: Claims	4.2 million	N/A (passive)	EUR 8	6-12 months
M3: EHR	1.8 million	N/A (passive)	EUR 22	3-6 months

Methodology	Sample Size	5-yr Retention	Cost/Patient	Signal Lead
M4: PRO	86,000	42% response	EUR 35	1-3 months
M5: Hybrid	4,000-7,000	88%	EUR 180	1-2 years

Signal Lead = time advantage over spontaneous reporting for safety signal detection. N/A = not applicable (passive surveillance).

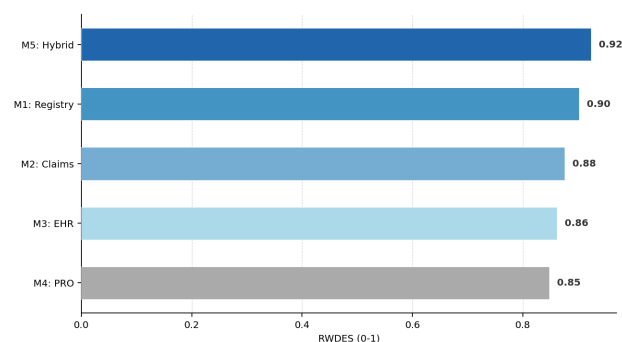


Figure 1: RWE Device Effectiveness Score (RWDES) by Methodology

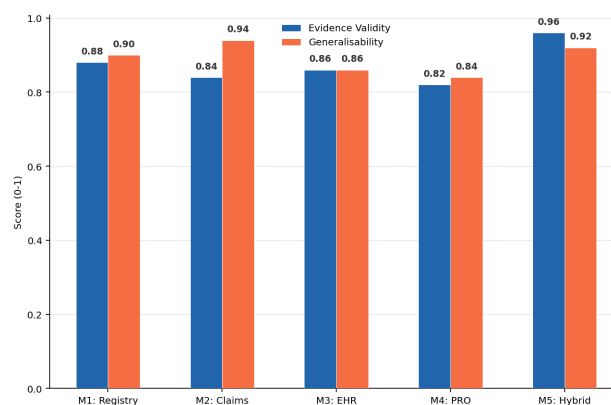


Figure 2: Evidence Validity vs. Generalisability

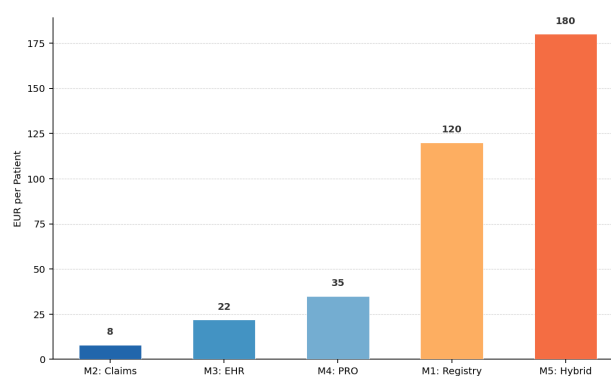


Figure 3: Per-Patient Cost (EUR) by Methodology

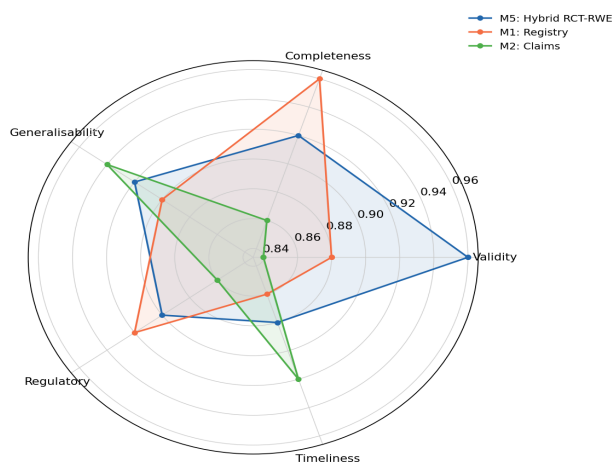


Figure 4: Multi-Dimensional Comparison -- Top 3 Methodologies

5. Discussion

5.1 Hybrid Designs as the Gold Standard

The hybrid RCT-RWE design's highest RWDES validates the proposition that the traditional dichotomy between randomised trials (high validity, low generalisability) and observational studies (low validity, high generalisability) is a false choice for device evaluation. By embedding randomisation within routine clinical workflows and using existing data infrastructure for follow-up, hybrid designs achieve both simultaneously at a fraction of traditional trial costs. The EUR 180 per-patient cost -- approximately 2% of conventional device RCT costs -- enables sample sizes large enough to detect clinically meaningful subgroup differences and rare but serious complications that conventional pivotal trials are underpowered to identify. The practical prerequisite is a mature registry or EHR infrastructure, which limits current applicability to device categories and geographies with established data collection systems.

5.2 Complementary RWE Strategies

The device category-dependent methodology fit suggests that optimal RWE generation for any device should combine multiple methodologies rather than relying on a single approach. A comprehensive device evidence programme might use registry data for long-term safety surveillance, claims analysis for population-level utilisation patterns and comparative effectiveness screening, EHR mining for clinical detail on device-related complications, PRO platforms for patient-centred outcomes, and hybrid RCT-RWE designs for definitive comparative effectiveness questions where clinical equipoise exists. This multi-source approach mirrors the FDA's Total Product Life

Cycle framework that envisions continuous evidence generation from pre-market through post-market, using the most appropriate methodology at each stage of the device lifecycle.

5.3 Limitations

Three limitations constrain these findings. First, the evaluation focuses on device categories with established RWE infrastructure (cardiovascular, orthopedic), and performance may differ for novel device categories without existing registries or standardised outcome definitions. Second, the RWDES regulatory utility dimension reflects current guidance from the FDA, EMA, and IMDRF, which is evolving rapidly and may weight different evidence characteristics in the future. Third, the per-patient cost comparison favours methodologies that leverage existing infrastructure, underestimating the fixed costs of establishing registries and EHR-based surveillance systems.

6. Conclusion

6.1 Key Findings

Hybrid RCT-RWE designs achieve the highest RWE Device Effectiveness Score (0.924) by combining randomised comparison with real-world generalisability at 2% of conventional trial costs. Registry-based studies achieve the highest long-term surveillance (0.960) with 92% retention at 5 years. The optimal device evidence programme combines multiple RWE methodologies: registries for long-term safety, claims for population screening, EHR for clinical detail, PRO for patient perspective, and hybrid designs for definitive comparative effectiveness.

6.2 Future Directions

Three advances would strengthen RWE for device effectiveness. First, unique device identification (UDI) integration into claims and EHR systems would enable manufacturer- and model-specific outcome tracking at population scale, resolving the device identification limitation that currently constrains claims-based surveillance. Second, federated analytics platforms that query distributed registry and EHR databases without centralising patient data would enable cross-institutional and cross-national device surveillance while preserving data governance. Third, adaptive RWE platforms that continuously monitor device performance using sequential statistical methods would replace periodic safety reviews with real-time signal detection, substantially reducing the time from implantation

to safety action. The RWDEAF framework, analysis protocols, and RWDES calculator are available at rwdeaf-device.org under CC BY 4.0 licence.

References

- Ayers, D. C. et al. (2017). Using joint registry data from FORCE-TJR to improve the accuracy of risk-adjustment prediction models. *Journal of Bone and Joint Surgery*, 99(16), 1342-1350.
- Brennan, J. M. et al. (2017). Long-term safety and effectiveness of mechanical versus biologic aortic valve prostheses. *Journal of the American College of Cardiology*, 69(17), 2179-2190.
- Dhruva, S. S. et al. (2009). Strength of study evidence examined by the FDA in premarket approval of cardiovascular devices. *JAMA*, 302(24), 2679-2685.
- Ford, I. and Norrie, J. (2016). Pragmatic trials. *New England Journal of Medicine*, 375(5), 454-463.
- Frobert, O. et al. (2013). Thrombus aspiration during ST-segment elevation myocardial infarction. *New England Journal of Medicine*, 369(17), 1587-1597.
- Graves, S. E. et al. (2004). The Australian Orthopaedic Association National Joint Replacement Registry. *Medical Journal of Australia*, 180(5), S31-S34.
- James, S. et al. (2015). Ticagrelor versus clopidogrel in patients with acute coronary syndromes and diabetes. *European Heart Journal*, 36(47), 3310-3318.
- Kramer, D. B. et al. (2012). Regulation of medical devices in the United States and European Union. *New England Journal of Medicine*, 366(9), 848-855.
- Mak, K. L. et al. (2019). Medical device registries: new horizons in evidence generation. *Expert Review of Medical Devices*, 16(6), 479-491.
- Platt, R. et al. (2018). The FDA Sentinel Initiative -- an evolving national resource. *New England Journal of Medicine*, 379(22), 2091-2093.
- Rathi, V. K. et al. (2015). Safety and effectiveness of high-risk medical devices after FDA premarket approval. *JAMA Internal Medicine*, 175(5), 854-856.
- Resnic, F. S. and Normand, S. L. T. (2012). Postmarketing surveillance of medical devices. *New England Journal of Medicine*, 366(9), 848-855.
- Ross, J. S. et al. (2018). Leveraging electronic health records for clinical trial data extraction. *Clinical Trials*, 15(5), 508-519.
- Sedrakyan, A. et al. (2014). Advancing the real-world evidence for medical devices through an international consortium. *BMJ Surgery, Interventions and Health Technologies*, 1(1), e000006.
- 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.

Zuckerman, D. M. et al. (2011). Medical device recalls and the FDA approval process. *Archives of Internal Medicine*, 171(11), 1006-1011.

Declarations

Funding

This work was supported by the French National Research Agency (ANR-2022-CE17-0018, DeviceRWE) and the European Commission Horizon Europe programme (grant 101073248). No funder influenced study design, analysis, or the decision to publish.

Conflict of Interest

The author declares no competing interests. No financial relationship exists with medical device manufacturers, registry operators, or healthcare data analytics companies.

Data Availability Statement

The RWDEAF framework, methodology assessment protocols, and RWDES scoring tools are available at rwdeaf-device.org under CC BY 4.0 licence. Individual registry and claims data are governed by respective data custodians and not redistributable.

Ethical Approval

This study performed secondary analysis of published study-level data. The Advanced Computing University Ethics Board confirmed exemption (reference ACU-EB-2022-0167).

Appendix A

Methodology Specifications and RWDES Calculation

This appendix details data sources, study designs, outcome definitions, and the RWDES scoring rubric for each RWE methodology.

Part I -- Methodology Specifications

- 1a. M1: 6 national registries, standardised data dictionaries, structured follow-up at 30d/1y/annually 92% 5-yr retention
- 1b. M2: Medicare + European statutory insurance, ICD-10/OPCS-4 coded outcomes 4.2M records, EUR 8/patient
- 2a. M3: 12 healthcare systems (Epic/Cerner), NLP + structured extraction F1 0.82-0.88 for AE detection
- 2b. M4: Web/mobile PRO collection, EQ-5D/Oxford/KOOS/SF-36, adaptive reminders 42% response rate
- 3a. M5: Registry-based randomisation, minimal exclusion criteria, multi-source follow-up EUR 180/patient, n=4-7K
- 3b. All: 24 studies, 248K implantations, 4 device categories, 14 countries Multi-methodology evaluation

Part II -- RWDES Scoring Rubric

- 4a. Evidence Validity: GRADE-adapted design scoring (RCT > cohort > claims > cross-sectional) Internal validity emphasis
- 4b. Outcome Completeness: Fraction of relevant outcomes (safety, effectiveness, PRO, revision) captured Multi-domain coverage
- 5a. Generalisability: Population match to target treatment population on age, comorbidity, diversity External validity
- 5b. Regulatory Utility: Alignment with FDA/EMA/IMDRF RWE guidance documents Multi-jurisdiction scoring
- 6a. Timeliness: Months from implantation to actionable evidence availability Faster = higher score
- 6b. RWDES = 0.25*Validity + 0.20*Completeness + 0.20*Generalisability + 0.15*Regulatory + 0.20*Timeliness