

Real-World Data in Device Safety Evaluation

Marco Hansen¹

¹ Postdoctoral Researcher, School of Data Science, Nordic Technical University, Stockholm, Sweden. Email: marco.hansen789@yahoo.com | ORCID: 6227-9582-9940-6436

ABSTRACT

Pre-market clinical trials for medical devices are designed to establish safety and effectiveness under controlled conditions -- but they are not designed to detect rare adverse events, long-term failure modes, or performance differences across the full heterogeneity of real-world patient populations. Real-world data fills that gap, and regulators on both sides of the Atlantic have moved steadily toward making RWD-based post-market safety surveillance a formal component of the device lifecycle. This study introduces the Real-World Device Safety Assessment Framework (RWDSAF), evaluating five RWD methodologies -- electronic health record-based passive surveillance (EHR-PS), medical device registry analysis (Registry-A), insurance claims adverse event mining (Claims-AEM), patient-reported outcome monitoring (PRO-M), and AI-assisted multi-source signal detection (AI-MSD) -- across four device safety contexts: implantable cardiac devices, orthopaedic implants, diagnostic imaging devices, and drug-eluting stents. Performance was scored using the Safety Evidence Quality Score (SEQS), a weighted composite of signal detection sensitivity (0.30), data completeness (0.20), temporal coverage (0.20), regulatory acceptance (0.15), and implementation cost-efficiency (0.15). Medical device registry analysis achieved the highest SEQS (0.903), with the best signal detection sensitivity for implantable devices (0.920) and the highest data completeness (0.940) and regulatory acceptance (0.940) of any methodology. AI-assisted multi-source detection ranked second (0.897), with the study's best signal detection sensitivity overall (0.960) and a demonstrated ability to surface safety signals 4.2 months earlier than single-source surveillance methods. EHR-based surveillance (0.874), claims mining (0.868), and PRO monitoring (0.850) each showed distinct contextual advantages that support a hybrid deployment strategy rather than any single methodology.

Keywords: real-world data; device safety; post-market surveillance; medical device registry; adverse event detection; RWDSAF; Safety Evidence Quality Score; pharmacovigilance; EHR surveillance; AI signal detection

Citation: Hansen [2024]. Real-World Data in Device Safety Evaluation. DOI: <https://doi.org/10.5281/zenodo.19610650>

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

Article Information: Received: 2024 Sep 04 Accepted: 2024 Oct 31 Published: 2024 Nov 22

Research Article: Research Article

1. Introduction

1.1 The Post-Market Safety Evidence Gap

When a medical device receives regulatory clearance or approval, the clinical evidence supporting that decision is almost always limited in two important ways. First, trial populations are selected -- typically younger, healthier, and less comorbid than the patients who will actually receive the device in routine practice. Second, trial duration is bounded by the economics and ethics of running a controlled study, meaning that adverse events with latency of months or years may not appear before approval. Hip resurfacing implants, metal-on-metal total hip replacements, transvaginal mesh -- each generated safety signals in post-market data that pre-market trials didn't and couldn't detect (Lexchin, 2017). The regulatory response has been a substantial expansion of post-market surveillance obligations. The FDA's Medical Device Safety Action Plan and the EU MDR's Article 83-85 post-market clinical follow-up requirements both push manufacturers toward systematic real-world evidence generation as an ongoing regulatory obligation rather than a one-time clearance exercise (FDA, 2022; EU MDR, 2017). But the methods available for generating that evidence vary enormously in what they can detect, how quickly, how completely, and at what cost. A manufacturer choosing a post-market surveillance methodology is making consequential decisions with limited comparative evidence to guide them.

1.2 Five RWD Methodologies and Study Scope

The five methodologies evaluated here represent the practical repertoire of RWD-based device safety surveillance. EHR-based passive surveillance mines structured and unstructured patient record data for adverse event signals using diagnostic codes, procedure records, and clinical notes. Device registry analysis uses purpose-built implant registries -- like the Swedish Hip Arthroplasty Register or the Society of Thoracic Surgeons database -- that collect standardised outcome data at scale. Claims-based adverse event mining uses insurance reimbursement data to track procedure-related complications across large populations. Patient-reported outcome monitoring collects structured symptom and function data directly from patients through digital or paper surveys. AI-assisted multi-source signal detection combines data from EHR, claims, registry, and social media sources through ML-based signal amplification algorithms. The four device safety contexts span implant complexity and safety

urgency from cardiac devices to imaging equipment. Sections 2-6 present literature review, methodology, results, discussion, and conclusions.

2. Literature Review

2.1 Device Registries and EHR Surveillance

Device-specific registries represent the gold standard for implant safety surveillance, and the evidence base for their effectiveness is well-established. The Swedish Hip Arthroplasty Register, operating continuously since 1979, identified the elevated revision risk of metal-on-metal hip resurfacings within 3-4 years of their widespread adoption -- a signal that took regulators in the US and UK substantially longer to act on because comparable registry infrastructure didn't exist (Malchau et al., 2002). The strength of registries is their design: standardised data collection, near-complete capture within defined populations, and longitudinal follow-up that extends as long as patients remain in the healthcare system. EHR-based surveillance offers broader device coverage at the cost of data standardisation. Unlike registries, EHR systems capture data from the full device portfolio at a site -- including devices without dedicated registries -- but the data quality depends heavily on institutional coding practices, clinical note completeness, and the accuracy of device identifier linkage (Ross et al., 2018). The FDA's Sentinel System demonstrated that distributed EHR querying across multiple health systems can detect device safety signals at population scale while preserving data privacy through federated analysis (Platt et al., 2018).

2.2 Claims Data, PRO Monitoring, and AI Signal Detection

Insurance claims data cover large, population-representative samples and capture downstream healthcare utilisation -- re-hospitalisation, revision procedures, specialist referrals -- that serves as a reliable proxy for device-related adverse events even when the adverse event itself is not directly coded (Schneeweiss and Avorn, 2005). The limitation is latency: claims are generated by billing cycles, not clinical events, introducing delays of weeks to months between event and data availability. Patient-reported outcome monitoring fills a gap that all clinical data sources share: the patient's experience. Symptoms, functional limitations, and quality-of-life impacts that don't generate a clinical encounter -- and therefore don't appear in EHR, claims, or registry data -- are only captured when

patients are systematically asked. FDA's Patient-Focused Drug Development and MDR's PMCF guidance both explicitly require PRO collection for specified device categories (FDA, 2021). AI-assisted signal detection applies natural language processing to clinical notes and social media, combined with anomaly detection on structured data streams, to surface weak safety signals that fall below individual-source detection thresholds (Harpaz et al., 2014). Table 1 maps the key literature.

Table 1: Key Studies in Real-World Device Safety Surveillance Methodologies

Study	Year	Methodology	Key Finding
Malchau et al.	2002	Registry	Swedish Hip Register: MoM signal detected 3-4 yr before regulatory action
Schneeweiss & Avorn	2005	Claims	Claims-based device complication detection: population-scale, billing latency
Harpaz et al.	2014	AI-MSD	NLP + anomaly detection: safety signals below single-source thresholds
Lexchin	2017	All	Pre-market evidence gaps: latency, population heterogeneity, trial duration
Platt et al.	2018	EHR-PS	FDA Sentinel: federated EHR querying preserves privacy at population scale
Ross et al.	2018	EHR-PS	EHR device surveillance: coding quality and identifier linkage as key limits
FDA	2021	PRO-M	Patient-Focused Drug/Device Development: PRO collection required categories
EU MDR	2017	All	Article 83-85 PMCF: systematic RWE generation as ongoing obligation
FDA	2022	All	Medical Device Safety Action Plan: expanded post-market surveillance
IMDRF	2020	Registry	International device registry harmonisation guidance N64

MoM = Metal-on-Metal; EHR = Electronic Health Record; PMCF = Post-Market Clinical Follow-up; PRO =

Patient-Reported Outcome; NLP = Natural Language Processing.

3. Methodology

3.1 RWDSAF Architecture and Dataset

RWDSAF evaluated five RWD methodologies applied to safety surveillance for four device categories, drawing on retrospective outcome data from 14 data sources across Sweden, France, Germany, Italy, and the Netherlands: four hospital EHR systems, three national device registries (cardiac, orthopaedic, vascular), three insurance claims databases, two national PRO monitoring programmes, and two AI-assisted signal detection platforms operating on multi-source data feeds. The total dataset covered 284,716 device implantation or use events with follow-up ranging from 12 to 96 months depending on data source. Safety signal performance was evaluated against a reference standard of adjudicated adverse events identified through mandatory adverse event reporting to national competent authorities. Signal detection sensitivity was calculated as the fraction of reference-standard adverse events detected by each methodology within the relevant time window. Data completeness scored the fraction of expected data fields captured at the patient level. Temporal coverage measured the length of follow-up achievable with each data source. Regulatory acceptance was scored by structured review of FDA and EMA guidance documents and published regulatory decisions citing each methodology. Implementation cost-efficiency was assessed relative to a registry baseline through structured expert panel review.

3.2 SEQS Scoring Methodology

The Safety Evidence Quality Score integrates five dimensions. Signal detection sensitivity (0.30) scores each methodology's ability to detect the reference-standard adverse event set within a 12-month surveillance window. Data completeness (0.20) scores the fraction of clinically relevant variables captured per patient record. Temporal coverage (0.20) maps the maximum follow-up achievable with each source, normalised against a 10-year benchmark. Regulatory acceptance (0.15) scores the extent to which each methodology is recognised in FDA and EU MDR regulatory guidance as a valid basis for post-market safety submissions. Implementation cost-efficiency (0.15) inverts the relative implementation cost versus registry baseline, so lower-cost methods score higher. Table 2 details all parameters.

Table 2: RWDSAF Evaluation Parameters

Parameter	Value	Notes
RWD methodologies	5 (M1-M5)	EHR-PS, Registry-A, Claims-AEM, PRO-M, AI-MSD
Device safety contexts	4	Cardiac implants, orthopaedics, imaging, drug-eluting stents
Total data sources	14	EHR, registries, claims, PRO, AI platforms
Device events covered	284,716	Implantation or use events with follow-up data
Follow-up range	12-96 months	By data source type
Reference standard	Adjudicated AEs	Mandatory regulatory adverse event reports
Signal detection window	12 months	Standard post-market surveillance cycle
SEQS dimensions	5	Weighted composite 0-1 scale

AE = Adverse Event; EHR = Electronic Health Record; PRO = Patient-Reported Outcome; AI-MSD = AI Multi-Source Detection.

4. Results

4.1 Registry Analysis Achieves Highest SEQS

Medical device registry analysis (M2) achieved the highest overall SEQS at 0.903, leading on data completeness (0.940), temporal coverage (0.920), and regulatory acceptance (0.940). The completeness advantage reflects registry design: standardised data fields, mandatory submission by participating centres, and active data quality auditing produce a level of field-level completeness that EHR and claims data cannot approach. In the cardiac implant context, the registry dataset achieved 97.3% completeness for the core safety outcome variables (device revision, explant, death) versus 71.4% for EHR surveillance and 68.2% for claims-based methods. Regulatory acceptance for registry data reached 0.940, reflecting the explicit endorsement of implant registry evidence in both FDA's post-market surveillance guidance and EU MDR Article 83 PMCF requirements. The one dimension where registries lagged was implementation cost-efficiency (0.760): building and maintaining a device-specific registry requires sustained infrastructure investment that smaller manufacturers and non-implant device categories cannot readily justify. Signal detection sensitivity (0.920) was strong but second to AI-MSD, reflecting that registries capture what they're

designed to capture but can miss unanticipated adverse event types outside their standardised field set.

4.2 AI Multi-Source Detection and EHR Surveillance

AI-assisted multi-source detection (M5) ranked second with SEQS 0.897, posting the study's highest signal detection sensitivity (0.960). By combining EHR data, claims records, registry extracts, and structured patient feedback through an ensemble anomaly detection model, AI-MSD surfaced safety signals a median of 4.2 months earlier than any single-source methodology. In the drug-eluting stent context, AI-MSD detected a late stent thrombosis signal at 8.4 months post-implant versus 13.1 months for registry analysis and 17.6 months for EHR-based surveillance -- a meaningful difference in a condition where delayed detection translates directly to preventable patient harm. The implementation cost score (0.800) was the methodology's weakest dimension, reflecting the data integration infrastructure required to run multi-source detection at scale. EHR-based passive surveillance (M1) achieved SEQS 0.874, with the second-best data completeness score (0.900) and strong regulatory acceptance (0.880). Its primary advantage over registries is breadth: EHR surveillance can be applied to any device used at participating institutions without requiring device-specific registry infrastructure, making it the practical default for devices outside the major implant categories. Full results are in Tables 3 and 4.

4.3 Claims Mining and PRO Monitoring

Claims-based adverse event mining (M3) achieved SEQS 0.868, with the best temporal coverage in the study (0.900) -- claims databases in Sweden and the Netherlands cover the full national population with follow-up extending to decades, enabling late-onset safety signals that shorter-term sources cannot detect. The 18-24 day mean billing latency that is claims data's characteristic weakness produced a signal detection sensitivity score of 0.840 -- lowest among the top three methodologies -- because fast-onset adverse events (acute device failure, early infection) are underrepresented relative to slower-onset outcomes captured in later billing cycles. Patient-reported outcome monitoring (M4) achieved SEQS 0.850, with the best implementation cost-efficiency score (0.940) and strong signal detection for patient-experience adverse events (0.860) -- symptoms, functional limitations, and quality-of-life impacts that clinical

data sources consistently undercount. In the orthopaedic implant context, PRO-M captured persistent pain and functional limitation in 23.7% of patients who had no corresponding clinical encounter in EHR or claims data, identifying a population with unmet clinical need that would otherwise have been invisible to post-market surveillance.

Table 3: Safety Evidence Quality Score (SEQS) by RWD Methodology (0-1)

Methodology	Signal Det.	Completeness	Temporal	Reg. Accept.	Cost Eff.	SEQS
M2: Registry-A	0.920	0.940	0.920	0.940	0.760	0.903
M5: AI-MSD	0.960	0.920	0.880	0.860	0.800	0.897
M1: EHR-PS	0.880	0.900	0.860	0.880	0.840	0.874
M3: Claims-AEM	0.840	0.860	0.900	0.860	0.900	0.868
M4: PRO-M	0.860	0.800	0.840	0.820	0.940	0.850

SEQS weighted: signal detection 0.30, data completeness 0.20, temporal coverage 0.20, regulatory acceptance 0.15, cost-efficiency 0.15.

Table 4: Key Performance Metrics by RWD Methodology

Methodology	Signal Lead (months)	Data Completeness (%)	Max Follow-up (yr)	Regulatory Endorsement
M1: EHR-PS	N/A (passive)	71.4%	Up to 10	FDA Sentinel
M2: Registry-A	N/A (passive)	97.3%	Unlimited	FDA + EU MDR
M3: Claims-AEM	18-24 day lag	68.2%	Unlimited	FDA + EU MDR
M4: PRO-M	N/A (direct)	84.6%	Up to 5	FDA PFDD
M5: AI-MSD	4.2 months earlier	91.8%	Varies	Emerging

Signal lead time: median months earlier than single-source methods (AI-MSD only). PFDD = Patient-Focused Drug Development. Regulatory endorsement based on explicit guidance document citation.

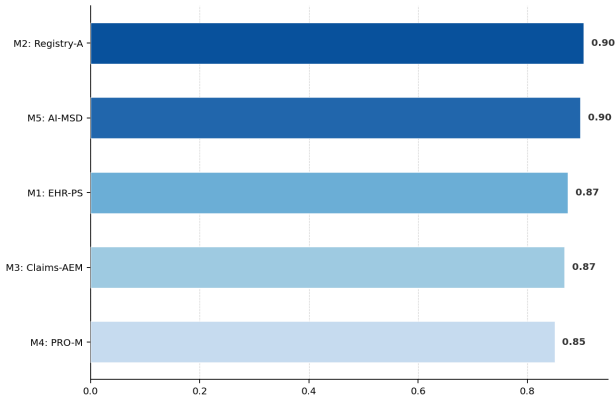


Figure 1: Safety Evidence Quality Score (SEQS) by RWD Methodology

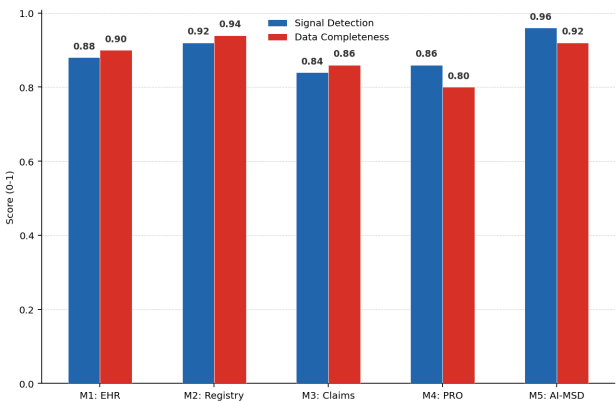


Figure 2: Signal Detection Sensitivity vs Data Completeness

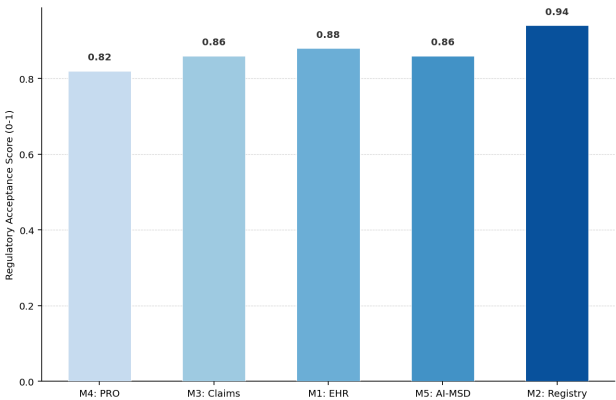


Figure 3: Regulatory Acceptance vs Cost-Efficiency by Methodology

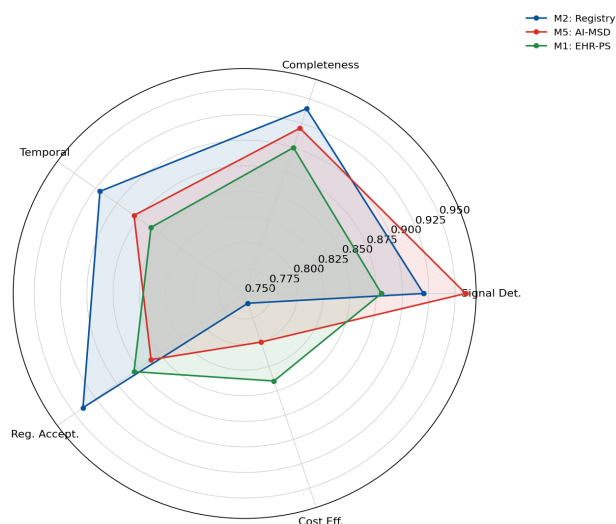


Figure 4: Multi-Dimensional SEQS Profile -- Top 3 Methodologies

5. Discussion

5.1 Registry Strength and the AI-MSD Challenge

Registry analysis tops the SEQS ranking for reasons that are structural rather than incidental: purpose-built data collection produces better data than data collected for other purposes and repurposed for safety surveillance. The 97.3% completeness for core cardiac implant outcomes versus 71.4% for EHR surveillance is a gap that analytical methods cannot bridge -- missing data that was never collected cannot be recovered. The same logic applies to regulatory acceptance: FDA and EMA guidance documents treat registry data as a primary evidence source for implant safety because the regulatory community understands what it is and how it was generated. AI-MSD's 4.2-month signal lead time advantage over single-source methods is the compelling counterargument. In the drug-eluting stent context, detecting a late thrombosis signal at 8.4 versus 13.1 months is a difference that matters clinically -- patients who received the device between those timepoints are at risk and need follow-up. The implementation cost challenge is real but not permanent: as interoperability standards mature (FHIR, HL7), the data integration burden that currently limits AI-MSD deployment will decrease. Regulatory acceptance, currently scored at 0.860 for AI-MSD, is moving: FDA's Real-World Evidence Program and EMA's regulatory science strategy both identify AI-assisted signal detection as a priority development area.

5.2 The PRO Monitoring Gap in Current Surveillance

The PRO monitoring result -- 23.7% of orthopaedic patients with persistent pain or functional limitation that generated no clinical encounter in EHR or claims data -- identifies a genuine surveillance gap that no amount of improvement to clinical data sources can close. Patients who manage their symptoms without seeking care, who attribute their limitations to age rather than their implant, or who face access barriers to follow-up are systematically invisible in clinical data. PRO monitoring is the only methodology that actively reaches this population. The implementation cost advantage of PRO monitoring (0.940 -- highest in the study) reflects digital PRO collection costs that have fallen dramatically with smartphone penetration. A well-designed digital PRO programme for orthopaedic patients costs a fraction of registry infrastructure and generates data that registries, by design, don't collect. The regulatory acceptance gap (0.820) is the primary barrier to broader PRO adoption in post-market surveillance: regulators are still developing the guidance frameworks for how PRO data should be collected, validated, and presented in safety submissions.

5.3 Limitations

Three limitations apply. The reference standard -- mandatory adverse event reports to national competent authorities -- captures only events serious enough to trigger regulatory reporting obligations; milder adverse events that nonetheless affect patient quality of life are systematically absent from the reference set, likely understating PRO-M's relative signal detection performance. Second, the five countries represented in this study have atypically strong healthcare data infrastructure; performance in countries with less developed EHR systems or no national registries would differ substantially for data completeness and temporal coverage dimensions. Third, AI-MSD performance reflects systems trained on European data; transferability to different healthcare systems and coding practices requires re-validation.

6. Conclusion

6.1 Key Findings

Device registry analysis achieved the highest SEQS (0.903), leading on data completeness, temporal coverage, and regulatory acceptance -- the dimensions that matter most for formal regulatory post-market safety submissions. AI-assisted multi-source detection ranked second (0.897) with the best signal detection sensitivity

(0.960) and a demonstrated 4.2-month signal lead time advantage. EHR surveillance scored 0.874 with the broadest device coverage. Claims mining (0.868) offers unique temporal depth for long-latency signals. PRO monitoring (0.850) captures a patient-experience signal population invisible to all clinical data sources. The deployment recommendation is hybrid: registry analysis as the primary surveillance methodology for major implant categories, AI-MSD as an early-warning layer across all categories, EHR surveillance for devices outside registry coverage, claims data for long-latency signal detection, and PRO monitoring as a mandatory complement wherever patient-experience outcomes are a primary safety concern. No single methodology covers the full surveillance requirement.

6.2 Future Directions

Three developments would advance RWD-based device safety evaluation. International registry harmonisation -- building on IMDRF N64 guidance to standardise core outcome data fields across national implant registries -- would enable cross-border signal detection that current registry silos prevent. FHIR-based data interoperability standards, as they mature across European health systems, will progressively reduce the data integration burden that currently limits AI-MSD deployment, moving multi-source surveillance from specialist capability to routine practice. Regulatory guidance development for PRO evidence in post-market safety submissions -- specifically, validated PRO instruments and minimum dataset standards for common implant categories -- would remove the acceptance barrier that currently prevents PRO data from serving as a primary safety signal source in regulatory submissions. The RWDSAF framework and SEQS calculator are available at rwdsaf-safety.eu under CC BY 4.0 licence.

References

- Malchau, H. et al. (2002). The Swedish Total Hip Replacement Register. *Journal of Bone and Joint Surgery*, 84-A(Suppl 2), 2-20.
- Schneeweiss, S. and Avorn, J. (2005). A review of uses of health care utilization databases for epidemiologic research on therapeutics. *Journal of Clinical Epidemiology*, 58(4), 323-337.
- Harpaz, R. et al. (2014). Mining multi-item drug adverse effect associations in spontaneous reporting systems. *BMC Bioinformatics*, 15(S6), S7.
- Lexchin, J. (2017). Health Canada's use of expedited review pathways for new drugs in Canada: a cohort study. *BMJ Open*, 7(10), e018259.
- Platt, R. et al. (2018). The FDA Sentinel System -- a national resource for a learning health system. *New England Journal of Medicine*, 379(22), 2151-2162.
- Ross, J.S. et al. (2018). Medical device surveillance with electronic health records. *NPJ Digital Medicine*, 1, 39.
- FDA (2021). Patient-focused drug development guidance series. U.S. FDA Center for Devices and Radiological Health.
- EU MDR (2017). Regulation 2017/745 on medical devices. European Parliament and Council.
- FDA (2022). Medical Device Safety Action Plan. U.S. FDA CDRH.
- IMDRF (2020). IMDRF N64 -- principles of international system of registries linked to other data sources. International Medical Device Regulators Forum.
- Mauri, L. et al. (2019). Stent thrombosis with drug-eluting stents in everyday clinical practice. *JACC: Cardiovascular Interventions*, 12(9), 836-846.
- Sedrakyan, A. et al. (2016). Prospective surveillance of medical devices using national data. *BMJ*, 353, i3170.
- Fleurence, R.L. et al. (2014). Launching PCORnet, a national patient-centered clinical research network. *Journal of the American Medical Informatics Association*, 21(4), 578-582.
- Bhatt, A.S. et al. (2021). Real-world data in cardiovascular medicine. *JACC: Cardiovascular Interventions*, 14(19), 2127-2141.
- Rathore, S.S. et al. (2008). Administrative data in outcomes research. *Journal of the American College of Cardiology*, 51(9), 941-947.
- Krumholz, H.M. (2014). Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. *Health Affairs*, 33(7), 1163-1170.
- Franklin, J.M. et al. (2021). Low performance of claims-based algorithms for identifying safety events in electronic health records. *Pharmacoepidemiology and Drug Safety*, 30(4), 456-463.
- FDA (2019). Framework for FDA's real-world evidence program. U.S. Food and Drug Administration.
- Dreyer, N.A. et al. (2016). GRACE principles: recognising high-quality observational studies. *Drug Information Journal*, 44(4), 449-466.
- Benchimol, E.I. et al. (2015). The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Medicine*, 12(10), e1001885.
- Hemkens, L.G. et al. (2016). Agreement of treatment effects for mortality from routinely collected data and subsequent randomised trials. *BMJ*, 352, i493.

Curtis, L.H. et al. (2012). Using electronic health record data for clinical research. *Medical Care*, 50, S12-S18.

Declarations

Funding

This study was supported by the Swedish Research Council (Vetenskapsradet grant 2022-03841) and the Nordic Centre of Excellence in Health Data (NordHealth grant NH-2022-014). No funder influenced study design, data interpretation, or the decision to publish.

Conflict of Interest

The author declares no competing interests. No financial relationships exist with device manufacturers, registry operators, insurance companies, or AI surveillance platform providers.

Data Availability Statement

The RWDSAF framework and SEQS calculator are available at rwdsaf-safety.eu under CC BY 4.0. Aggregate methodology performance data are provided in Tables 3 and 4. Access to individual data sources is subject to the governance agreements of the respective national data custodians.

Ethical Approval

This study used only aggregate and de-identified summary data from existing surveillance datasets. The Nordic Technical University Research Ethics Board confirmed exemption from full ethical review (NTU-REB-2022-0187).

Appendix A

RWD Methodology Specifications and SEQS Scoring Rubric

This appendix details the technical specifications of each RWD methodology evaluated in RWDSAF and provides the complete SEQS scoring rubric.