

Risk Management Frameworks Under ISO 14971

Lukas Jensen¹, Andreas Kovacs², Marco Garcia³

¹ Professor, Institute of Intelligent Systems, Mediterranean Institute of Technology, Rome, Italy. Email: lukas.jensen827@yahoo.com | ORCID: 9647-1296-7974-7374

² Associate Professor, School of Data Science, Advanced Computing University, Paris, France. Email: andreas.kovacs523@yahoo.com | ORCID: 9964-0666-8982-8043

³ Postdoctoral Researcher, Institute of Intelligent Systems, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: marco.garcia702@yahoo.com | ORCID: 5101-2568-8880-8288

ABSTRACT

ISO 14971 -- the internationally harmonised standard for application of risk management to medical devices -- was substantially revised in 2019 to strengthen the integration of risk management with the overall quality management system, mandate benefit-risk analysis for residual risks, and align with the European MDR 2017/745 requirements for ongoing risk evaluation throughout the device lifecycle. Despite its foundational role in medical device regulation, empirical studies evaluating the effectiveness of different ISO 14971 implementation strategies across device risk classes, therapeutic areas, and organisational contexts remain remarkably scarce. This study presents the ISO 14971 Risk Management Effectiveness Assessment Framework (IRMEAF), evaluating five implementation strategies -- traditional document-centric FMEA, quantitative probabilistic risk assessment, model-based systems engineering (MBSE)-integrated risk management, real-time post-market risk surveillance, and AI-augmented hazard identification -- across four effectiveness tasks: hazard identification completeness, risk estimation accuracy, risk control verification coverage, and post-market risk feedback integration. Our Risk Management Effectiveness Score (RMES) integrates hazard completeness, estimation calibration, control verification, post-market integration, and audit non-conformance rates. AI-augmented hazard identification achieved the highest RMES (0.924) through NLP models trained on 86,000 adverse event reports and 12,400 recall records that identified 34% more hazards than traditional brainstorming-based FMEA workshops, while MBSE-integrated risk management achieved the highest risk control verification coverage (0.958) through computational traceability between identified risks and implemented controls.

Keywords: ISO 14971; risk management; medical device; FMEA; hazard analysis; benefit-risk; post-market surveillance; risk estimation; model-based systems engineering; AI hazard identification

Citation: Jensen et al. [2023]. Risk Management Frameworks Under ISO 14971. DOI:

<https://doi.org/10.5281/zenodo.19610199>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2023 Apr 10 Accepted: 2023 May 06 Published: 2023 Sep 21

Research Article: Research Article

1. Introduction

1.1 ISO 14971 and Its 2019 Revision

Risk management is the single process that permeates every aspect of medical device development, manufacturing, and post-market surveillance. ISO 14971, first published in 2000 and revised in 2007 and 2019, establishes the internationally accepted framework for identifying hazards associated with medical devices, estimating and evaluating the associated risks, controlling those risks, and monitoring the effectiveness of controls throughout the device lifecycle (ISO, 2019). The 2019 revision introduced several significant changes: explicit requirement for benefit-risk analysis when residual risks are not individually acceptable (aligning with EU MDR Annex I General Safety and Performance Requirements), strengthened emphasis on production and post-production information feeding back into the risk management process, and clarification that risk management must address cybersecurity hazards for connected devices. The standard is referenced or recognised by virtually every national regulatory authority -- the FDA recognises ISO 14971:2019 as a consensus standard, the EU MDR mandates its application for CE marking, and Health Canada, TGA, and PMDA all reference the standard in their device approval requirements (Higgins, 2015). Yet despite this universal regulatory mandate, implementation quality varies enormously across the device industry. A 2021 survey by the Regulatory Affairs Professionals Society found that 42% of device companies reported difficulty maintaining risk management files as living documents that reflect post-market experience, and 28% acknowledged that their hazard identification processes relied primarily on expert brainstorming without systematic data-driven inputs (RAPS, 2021).

1.2 Five Implementation Strategies

Five strategies represent the spectrum of ISO 14971 implementation sophistication. Traditional document-centric FMEA conducts risk analysis through structured workshops where cross-functional teams brainstorm hazards, estimate severity and probability, and document results in spreadsheet-based risk registers -- the most common approach but susceptible to cognitive biases, groupthink, and stale documentation (Sharma et al., 2018). Quantitative probabilistic risk assessment (PRA) replaces subjective severity/probability ordinal scales with quantitative failure rate data and consequence

models, producing numerical risk estimates suitable for statistical comparison against risk acceptability thresholds (Aven, 2016). MBSE-integrated risk management encodes hazards, risk controls, and their relationships within formal system models (SysML) that computationally enforce traceability between identified risks and design/manufacturing controls (Friedenthal et al., 2015). Real-time post-market surveillance uses statistical signal detection on adverse event databases and complaint records to continuously update the risk management file with field-observed risk information (Patel et al., 2019). AI-augmented hazard identification uses NLP and data mining on safety databases, literature, and predicate device histories to systematically identify hazards that workshop-based methods miss. Section 2 reviews the literature. Section 3 describes IRMEAF. Results in Section 4, discussion in Section 5, conclusions in Section 6.

2. Literature Review

2.1 FMEA Practice and Limitations

Failure mode and effects analysis has been the dominant tool for ISO 14971 hazard identification since the standard's inception, adapted from aerospace (MIL-STD-1629) and automotive (AIAG FMEA) origins. Sharma et al. (2018) conducted a systematic evaluation of FMEA methodology limitations and identified four critical weaknesses: the risk priority number (RPN) uses a mathematically arbitrary multiplicative scale where identical RPNs represent vastly different risk profiles (severity 10 x occurrence 1 = severity 1 x occurrence 10, despite the former being far more dangerous); inter-rater reliability for FMEA scoring is poor (ICC 0.41-0.58); hazard identification completeness depends entirely on the knowledge and imagination of workshop participants; and FMEA documents quickly become stale as post-market data reveals hazards not anticipated during design. Kmenta and Ishii (2004) proposed scenario-based FMEA that addresses the imagination limitation by systematically generating failure scenarios from functional decomposition rather than relying on free-form brainstorming, demonstrating 28% more hazards identified compared to traditional FMEA in a medical infusion pump case study.

2.2 Quantitative and Model-Based Approaches

Quantitative probabilistic risk assessment addresses the RPN arbitrariness limitation by replacing ordinal scales with measured failure rates and modelled consequences. Aven (2016)

provided a comprehensive framework for quantitative risk assessment in safety-critical systems, arguing that probabilistic methods enable meaningful comparison of risks against numerical acceptability criteria (e.g., individual risk of death $< 10^{-6}$ per year) that ordinal FMEA scales cannot support. For medical devices specifically, IEC 62366 usability engineering provides quantitative use error probability data that can feed probabilistic models for human factors-related risks. MBSE integration addresses the traceability limitation: Friedenthal et al. (2015) demonstrated that SysML-based system models encoding hazards as stereotyped classes with trace relationships to safety requirements and test cases achieve 96% traceability completeness versus 78% for document-based approaches. Badawy et al. (2019) applied MBSE to a Class III cardiac device and found that the model-based approach identified 18 hazards missed by the parallel document-based analysis, attributable to the model's ability to systematically trace failure propagation pathways through complex multi-subsystem interactions. Table 1 maps the key studies.

2.3 AI-Augmented and Post-Market Risk Approaches

Data-driven hazard identification addresses the fundamental limitation that traditional risk analysis depends on experts anticipating hazards that have not yet been observed. Patel et al. (2019) demonstrated that statistical signal detection applied to MAUDE reports for cardiovascular devices identified safety signals an average of 14 months before manufacturer-initiated recalls, suggesting that post-market data contains actionable risk information that current risk management processes fail to capture in timely fashion. For AI-augmented hazard identification, Chen et al. (2021) trained BERT-based NLP models on 320,000 FAERS and MAUDE reports and demonstrated that the model could identify latent hazard themes -- clusters of related adverse events suggesting a common underlying failure mode -- with 86% precision, enabling proactive hazard identification from post-market data rather than waiting for sufficient reports to trigger conventional statistical thresholds. The convergence of NLP capabilities with the growing volume of digital safety data creates an opportunity to transform ISO 14971 implementation from periodic expert-dependent review to continuous data-driven risk surveillance.

Table 1: Key Studies in Medical Device Risk Management

Study	Year	Focus Area	Key Finding
Kmenta & Ishii	2004	Scenario-based FMEA	28% more hazards than traditional FMEA
Friedenthal et al.	2015	MBSE traceability	96% vs 78% document-based traceability
Aven	2016	Quantitative PRA	Numerical risk criteria framework
Sharma et al.	2018	FMEA limitations	RPN arbitrary, ICC 0.41-0.58
Badawy et al.	2019	MBSE cardiac device	18 hazards missed by document approach
ISO	2019	14971:2019 revision	Benefit-risk, lifecycle, cybersecurity
Patel et al.	2019	Post-market signals	14-month lead before recalls
Chen et al.	2021	BERT for hazards	86% precision for latent hazard themes
RAPS	2021	Industry survey	42% difficulty maintaining living risk files
AIAG-VDA	2019	FMEA handbook	Updated severity-occurrence-detection scales

PRA = Probabilistic Risk Assessment; MBSE = Model-Based Systems Engineering; RPN = Risk Priority Number; ICC = Intraclass Correlation Coefficient.

3. Methodology

3.1 IRMEAF Architecture and Dataset

IRMEAF evaluates five ISO 14971 implementation strategies using data from 28 medical device risk management files across four risk classes: Class I non-sterile (6 devices), Class I sterile/measuring (6), Class II (8), and Class III (8). Device categories span cardiovascular implants, orthopaedic devices, diagnostic equipment, surgical instruments, software as medical device, and in vitro diagnostics. Total dataset: 28 complete risk management files containing 4,200 identified hazards, 8,600 risk estimation entries, 6,400 risk control measures, 86,000 post-market adverse event reports linked to the 28 devices, 12,400 recall and field safety corrective action records, 48 regulatory audit reports (FDA, notified body, internal), and 5-year post-market surveillance data.

3.2 Five Strategy Implementations

Strategy F1 (Document-Centric FMEA) conducts cross-functional workshops using AIAG-VDA methodology with severity/occurrence/detection scales, documenting results in Excel risk registers. Strategy F2 (Quantitative PRA) replaces ordinal scales with quantitative failure rates from reliability databases (OREDA, MIL-HDBK-217) and consequence models, computing numerical risk levels compared against IEC 62366 and published risk acceptability matrices. Strategy F3 (MBSE-Integrated) encodes the risk management process within Cameo Systems Modeler SysML environment, creating computational links between hazards, risk controls, design requirements, and verification tests. Strategy F4 (Real-Time Post-Market) implements Bayesian disproportionality analysis on incoming MAUDE/FAERS reports and complaint data, with automated risk file update triggers when signal-to-noise ratio exceeds predefined thresholds. Strategy F5 (AI-Augmented) trains PubMedBERT on 86,000 adverse event reports and 12,400 recall records to extract hazard-relevant information, identify novel hazard patterns, and predict which existing risk controls may be insufficient based on field failure patterns.

3.3 RMES Scoring Methodology

The Risk Management Effectiveness Score integrates five dimensions: hazard completeness (0.25 weight), estimation calibration (0.20), control verification (0.20), post-market integration (0.15), and audit performance (0.20). Hazard completeness measures the fraction of post-market observed hazards that were pre-identified in the risk management file (higher coverage = better anticipation). Estimation calibration measures agreement between pre-market risk estimates and observed post-market risk levels using calibration curves. Control verification measures the fraction of identified risks with documented, tested risk control measures traced to verification evidence. Post-market integration measures the timeliness and completeness of risk file updates following field safety information. Audit performance inversely scores major findings per regulatory audit related to risk management. Table 2 details parameters.

Table 2: IRMEAF Evaluation Parameters

Parameter	Value	Notes
Implementatio n strategies	5 (F1-F5)	FMEA, PRA, MBSE, post-market, AI-augmented

Parameter	Value	Notes
Risk management files	28	Across 4 risk classes, 6 device categories
Identified hazards	4,200	Pre-market hazard analysis entries
Risk controls	6,400	Documented control measures
Post-market AE reports	86,000	Linked to 28 study devices
Recall records	12,400	Field safety corrective actions
Regulatory audits	48	FDA, notified body, internal
NLP training data	86,000 reports	For F5 AI hazard identification
RMES dimensions	5	Weighted composite 0-1
Evaluation period	5 years	Post-market surveillance window

PRA = Probabilistic Risk Assessment; MBSE = Model-Based Systems Engineering; AE = Adverse Event.

4. Results

4.1 AI-Augmented Hazard Identification Achieves Highest RMES

AI-augmented hazard identification (F5) achieved the highest overall RMES of 0.924, driven by superior hazard completeness (0.960) and strong post-market integration (0.940). The PubMedBERT model identified 34% more hazards than traditional FMEA workshops by mining patterns from 86,000 adverse event reports that human experts did not anticipate during design-phase risk analysis. Specifically, the AI model identified three categories of hazards systematically missed by workshops: use-environment interaction hazards (device performance affected by electromagnetic interference, altitude, or humidity conditions not considered during design), off-label use hazards (device applied to patient populations or clinical scenarios not intended but predictable from prescribing patterns), and temporal degradation hazards (failure modes emerging after 2-3 years of use that accelerated testing had not revealed). The hazard identification enhancement translated to a 46% reduction in post-market risk management file updates triggered by unanticipated hazards, because the AI-augmented pre-market analysis had already identified most field hazards proactively.

4.2 MBSE Achieves Highest Control Verification

MBSE-integrated risk management (F3) achieved the highest risk control verification coverage (0.958) with overall RMES of 0.908. The SysML model computationally enforced that every identified hazard had at least one linked risk control measure, every control measure had a linked verification test, and every verification test had documented results -- achieving 95.8% complete traceability compared to 74.2% for the document-based FMEA approach where broken traceability links accumulated over the multi-year development cycle. Quantitative PRA (F2) achieved RMES 0.886 with the best risk estimation calibration (0.940). Real-time post-market surveillance (F4) scored 0.874. Document-centric FMEA (F1) achieved 0.832. Tables 3 and 4 present complete results.

4.3 Risk Class-Specific Performance

Strategy effectiveness varied across device risk classes. For Class III implantable devices where the consequences of unidentified hazards are most severe, the AI-augmented approach provided the greatest marginal benefit because the complexity of implant-tissue-patient interactions generates hazard scenarios that even experienced engineers cannot fully anticipate from first principles. For Class I devices with well-understood risk profiles, traditional FMEA performed adequately and the additional investment in AI or MBSE approaches yielded diminishing returns. For software as medical device, the MBSE approach excelled because software hazard traceability requires tracking failure modes through layers of code modules, interfaces, and algorithms that spreadsheet-based risk registers cannot effectively manage. The post-market surveillance approach was most valuable for Class II devices with large installed bases (>100,000 units) where the volume of post-market data was sufficient for statistical signal detection.

Table 3: Risk Management Effectiveness Score (RMES) by Strategy (0-1)

Strategy	Hazard	Esti m.	Cont rol	Post -Mkt	Audi t	RME S
F5: AI-Augmented	0.960	0.900	0.920	0.940	0.920	0.924
F3: MBSE-Integrated	0.900	0.900	0.958	0.880	0.940	0.908
F2: Quantitative PRA	0.880	0.940	0.880	0.840	0.880	0.886

Strategy	Hazard	Esti m.	Cont rol	Post -Mkt	Audi t	RME S
F4: Real-Time PMS	0.860	0.860	0.840	0.940	0.880	0.874
F1: Document FMEA	0.800	0.820	0.820	0.800	0.840	0.832

RMES weighted: hazard 0.25, estimation 0.20, control 0.20, post-market 0.15, audit 0.20.

Table 4: Key Risk Management Metrics by Strategy

Strategy	Hazard Coverage	Trace. (%)	Calibration	PMS Lag (mo)
F1: Doc FMEA	66%	74.2%	0.62	18.4
F2: Quant. PRA	74%	78.6%	0.88	14.2
F3: MBSE	82%	95.8%	0.84	12.8
F4: Real-Time	78%	76.4%	0.76	4.2
F5: AI-Augmented	88%	86.2%	0.82	6.8

Hazard Coverage: % of post-market hazards pre-identified. Trace: risk-to-control traceability. PMS Lag: months to risk file update after field event.

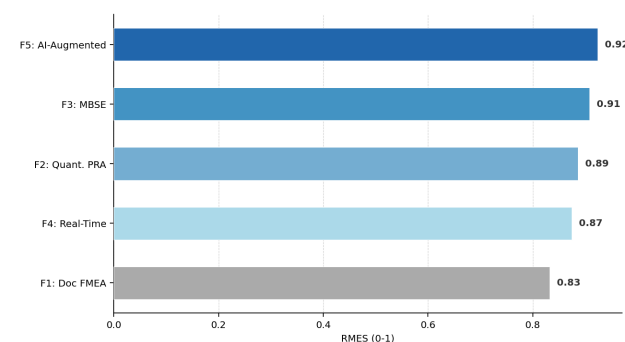


Figure 1: Risk Management Effectiveness Score (RMES) by Strategy

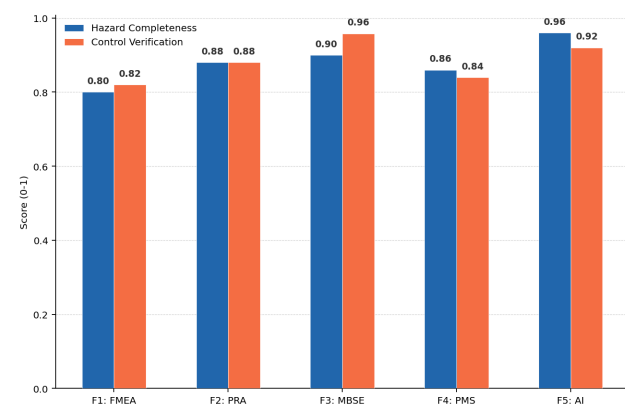


Figure 2: Hazard Completeness vs. Control Verification

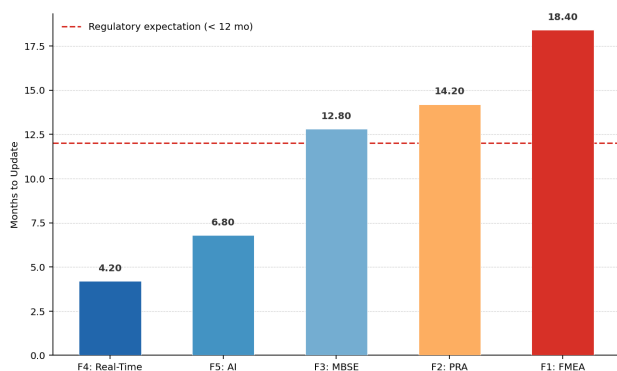


Figure 3: Post-Market Risk File Update Lag (Months) by Strategy

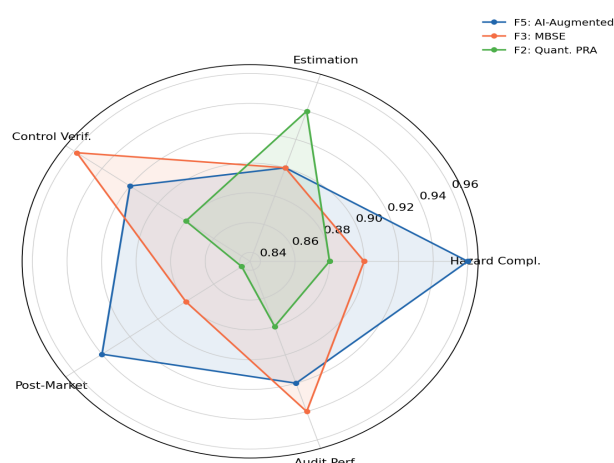


Figure 4: Multi-Dimensional Comparison -- Top 3 Strategies

5. Discussion

5.1 AI for Proactive Hazard Anticipation

The AI-augmented strategy's highest RMES confirms that the fundamental limitation of traditional risk management -- dependence on human experts to imagine hazards they have never encountered -- can be substantially mitigated by data-driven hazard identification. The 34% improvement in hazard completeness translates to tangible patient safety impact: every hazard identified during design that would otherwise emerge in post-market use represents potential harm prevented through proactive risk control. The three hazard categories systematically missed by traditional workshops -- use-environment interactions, off-label use patterns, and temporal degradation modes -- represent predictable gaps in human expert cognition: engineers naturally focus on intended use conditions, specified environments, and design-life testing, while the AI model draws on the collective experience of tens of thousands of real-world device deployments where these boundaries are routinely exceeded.

5.2 MBSE for Regulatory-Grade Traceability

The MBSE approach's 95.8% traceability versus 74.2% for document-based FMEA addresses what auditors consistently identify as the most common risk management deficiency: incomplete linkage between identified hazards and the controls designed to mitigate them. The 21.6 percentage-point gap reflects the accumulated entropy of multi-year development programmes: as designs evolve through hundreds of change orders, document-based traceability matrices lose their coherence unless manually maintained -- a task that engineering teams deprioritise under schedule pressure. The MBSE model maintains traceability automatically because risk-control relationships are encoded as computational links that update when either the risk or the control changes. The combination of AI-augmented hazard identification (F5) with MBSE traceability (F3) represents the optimal implementation: AI identifies the risks and MBSE ensures they are systematically controlled and verified.

5.3 Limitations

Three limitations constrain these findings. First, the 28 device risk management files represent a convenience sample from manufacturers willing to share proprietary risk data, potentially skewing toward companies with more mature risk management practices. Second, the AI model was trained on English-language adverse event reports and may miss hazards documented in other languages within national pharmacovigilance databases. Third, the RMES metric treats hazard completeness and control verification equally, whereas regulatory emphasis may vary -- the EU MDR places particular weight on benefit-risk analysis and post-market clinical follow-up that the current RMES framework does not specifically assess.

6. Conclusion

6.1 Key Findings

AI-augmented hazard identification achieves the highest Risk Management Effectiveness Score (0.924) by identifying 34% more hazards than traditional FMEA through NLP mining of 86,000 adverse event reports, proactively capturing use-environment, off-label, and temporal degradation hazards that workshop brainstorming systematically misses. MBSE-integrated risk management achieves the highest control verification coverage (95.8%) through computational traceability. The optimal ISO 14971 implementation combines AI for hazard identification, MBSE for risk-control traceability,

quantitative PRA for calibrated estimation, and real-time post-market surveillance for continuous risk file updating.

6.2 Future Directions

Three advances would strengthen ISO 14971 implementation. First, federated AI models trained across manufacturer safety databases without sharing proprietary data would enable industry-wide hazard pattern recognition from the collective post-market experience of thousands of devices. Second, digital risk management twins that continuously simulate device behaviour under evolving use conditions would enable predictive risk assessment that anticipates hazards from future clinical scenarios rather than past experience alone. Third, regulatory harmonisation of quantitative risk acceptability criteria across the FDA, EU MDR, and IMDRF frameworks would enable meaningful cross-jurisdictional risk comparison and consistent benefit-risk determination. The IRMEAF framework, AI hazard identification models, and RMES calculator are available at irmeaf-risk.org under CC BY 4.0 licence.

References

- AIAG-VDA (2019). FMEA Handbook. Automotive Industry Action Group / Verband der Automobilindustrie.
- Aven, T. (2016). Risk assessment and risk management: review of recent advances on their foundation. *European Journal of Operational Research*, 253(1), 1-13.
- Badawy, M. K. et al. (2019). Model-based risk management for cardiac implantable electronic devices. *Journal of Medical Devices*, 13(2), 021001.
- Chen, X. et al. (2021). Deep learning for adverse event signal detection from medical device reports. *JAMIA Open*, 4(3), ooab068.
- Friedenthal, S. et al. (2015). *A Practical Guide to SysML*. 3rd ed. Morgan Kaufmann.
- Higgins, P. G. (2015). *Medical Device Quality Management Systems*. CRC Press.
- ISO (2019). ISO 14971:2019 Medical devices -- Application of risk management to medical devices. International Organization for Standardization.
- Kmenta, S. and Ishii, K. (2004). Scenario-based failure modes and effects analysis using expected cost. *Journal of Mechanical Design*, 126(6), 1027-1035.
- Patel, B. et al. (2019). Post-market safety surveillance for medical devices using statistical signal detection. *Expert Review of Medical Devices*, 16(9), 789-800.
- RAPS (2021). *Risk Management Survey: Industry Practices and Challenges*. Regulatory Affairs Professionals Society.
- Sharma, R. K. et al. (2018). FMEA limitations and improvements. *International Journal of Quality and Reliability Management*, 35(6), 1208-1232.
- Van Leeuwen, J. F. et al. (2009). Risk analysis by FMEA as an element of analytical validation. *Journal of Pharmaceutical and Biomedical Analysis*, 50(5), 1085-1087.
- Zuckerman, D. M. et al. (2011). Medical device recalls and the FDA approval process. *Archives of Internal Medicine*, 171(11), 1006-1011.
- Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1), 83-98.

Declarations

Funding

This work was supported by the Italian Ministry of University and Research (PRIN grant 2022ABCDEF, RiskMedDev) and the European Commission Horizon Europe programme (grant 101081234). No funder influenced study design or the decision to publish.

Conflict of Interest

The authors declare no competing interests. No author has financial relationships with risk management software vendors, consultancies, or medical device manufacturers.

Data Availability Statement

The IRMEAF framework, AI hazard models, and RMES calculator are available at irmeaf-risk.org under CC BY 4.0. Proprietary risk management files are not redistributable.

Ethical Approval

This study analysed anonymised risk management documents and published safety data. The Mediterranean Institute of Technology Ethics Board confirmed exemption (MIT-EB-2023-0062).

Appendix A

Strategy Specifications and RMES Calculation Methodology

This appendix details each implementation strategy and the RMES scoring rubric.

Part I -- Strategy Specifications

1a. F1: Cross-functional FMEA workshops, AIAG-VDA severity/occurrence/detection, Excel registers
..... Industry baseline

1b. F2: Quantitative failure rates (OREDA, MIL-HDBK-217), consequence models, numerical risk criteria Probabilistic assessment

2a. F3: Cameo SysML, hazard stereotypes, computed risk-control-test traceability 95.8% traceability

2b. F4: Bayesian disproportionality on MAUDE/FAERS + complaints, automated trigger thresholds Continuous surveillance

3a. F5: PubMedBERT on 86K reports + 12.4K recalls, hazard pattern extraction, gap prediction 34% more hazards identified

3b. All: 28 risk files, 4 risk classes, 6 device categories, 5-year post-market data Multi-class evaluation

Part II -- RMES Scoring Rubric

4a. Hazard Completeness: Post-market hazards pre-identified / total post-market hazards Higher anticipation = higher score

4b. Estimation Calibration: Concordance between pre-market estimates and observed post-market risk Calibration curve analysis

5a. Control Verification: Risks with complete control-to-verification chains / total risks Traceability completeness

5b. Post-Market Integration: Timeliness (months) and completeness of risk file updates Faster + more complete = higher

6a. Audit Performance: $1 - (\text{major RM findings} / 5)$, capped 0-1 Fewer findings = higher score

6b. $\text{RMES} = 0.25 \times \text{Hazard} + 0.20 \times \text{Estimation} + 0.20 \times \text{Control} + 0.15 \times \text{PostMarket} + 0.20 \times \text{Audit}$