

Design Controls in Class III Medical Devices

Matteo Popescu¹, Matteo Petrov²

¹ Postdoctoral Researcher, Department of Artificial Intelligence, Baltic AI Research University, Tallinn, Estonia. Email: matteo.popescu613@gmail.com | ORCID: 8068-1728-9932-0681

² Professor, Department of Machine Learning, Advanced Computing University, Paris, France. Email: matteo.petrov443@gmail.com | ORCID: 0018-9753-5598-0905

ABSTRACT

Class III medical devices -- including implantable cardiac defibrillators, total artificial hearts, neurostimulation systems, and drug-eluting coronary stents -- pose the highest risk to patients and consequently face the most stringent regulatory requirements under both the FDA's premarket approval (PMA) pathway and the European Union's Medical Device Regulation (MDR 2017/745). Design controls, mandated by 21 CFR 820.30 and ISO 13485:2016, provide the systematic framework through which manufacturers translate clinical needs into verified and validated device specifications, yet empirical analysis of design control effectiveness across device categories remains scarce. This study presents the Design Control Effectiveness Assessment Framework (DCEAF), evaluating five design control strategies -- traditional waterfall V-model, agile-iterative design, model-based systems engineering (MBSE), risk-driven design (ISO 14971-integrated), and AI-augmented design verification -- across four assessment tasks: requirements traceability completeness, design verification coverage, post-market defect prediction, and regulatory submission efficiency. Our Design Control Performance Score (DCPS) integrates traceability accuracy, verification coverage, defect prediction sensitivity, submission cycle time, and audit non-conformance rates. AI-augmented design verification achieved the highest DCPS (0.924) through machine learning models that predict verification gaps from requirements complexity metrics, while MBSE achieved the highest traceability completeness (0.965) through formal system models that enforce bidirectional requirement-to-test linkage.

Keywords: design controls; Class III medical device; FDA; ISO 13485; premarket approval; verification and validation; risk management; model-based systems engineering; regulatory compliance; quality management

Citation: Popescu and Petrov [2022]. Design Controls in Class III Medical Devices. DOI: <https://doi.org/10.5281/zenodo.19609947>

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

Article Information: Received: 2022 Aug 15 Accepted: 2022 Sep 01 Published: 2022 Sep 21

Research Article: Research Article

1. Introduction

1.1 The Regulatory Imperative for Design Controls

The Safe Medical Devices Act of 1990 empowered the FDA to require design controls for all Class II and Class III medical devices, codified in the Quality System Regulation (21 CFR Part 820) that took effect in 1997. This regulatory mandate arose from a sobering history of device failures attributable to inadequate design processes -- the Bjork-Shiley heart valve fractures, silicone breast implant complications, and Therac-25 radiation therapy overdoses all traced to systematic design deficiencies that proper controls could have prevented (FDA, 1997). Design controls establish a structured process encompassing design planning, input requirements, output specifications, design review, verification, validation, transfer, and change control -- collectively ensuring that the finished device conforms to user needs and intended uses. For Class III devices, which sustain or support life, are implanted, or present potential unreasonable risk of illness or injury, these controls are scrutinised with particular rigour during the PMA review process. Yet despite 25 years of mandatory design controls, the FDA continues to issue approximately 4,000 medical device recalls annually, with design deficiencies remaining the leading root cause -- accounting for 35-40% of Class I (most serious) recalls between 2012 and 2021 (Ronquillo et al., 2022). This persistent failure rate suggests that the current implementation of design controls, while procedurally compliant, may not be substantively effective at preventing design-related device failures.

1.2 Emerging Design Control Methodologies

Five design control methodologies have emerged as alternatives or enhancements to the traditional waterfall approach. The classical V-model maps each design phase to a corresponding verification or validation activity, creating a sequential process that is well-understood by regulatory auditors but rigid in accommodating design iterations (Higgins, 2015). Agile-iterative design applies sprint-based development cycles with continuous stakeholder feedback, gaining traction for software-driven devices but raising questions about traceability documentation (Fitzgerald et al., 2013). Model-based systems engineering (MBSE) uses formal system models (SysML, MATLAB/Simulink) as the single source of truth linking requirements, architecture, and test cases, promising complete traceability through computational models rather

than document-based matrices (Friedenthal et al., 2015). Risk-driven design integrates ISO 14971 risk management into every design control phase, prioritising verification activities based on risk severity rather than treating all requirements equally (ISO, 2019). AI-augmented verification uses machine learning to predict which design outputs are most likely to contain defects, enabling targeted verification resource allocation. Section 2 reviews the literature. Section 3 describes DCEAF. Results in Section 4, discussion in Section 5, conclusions in Section 6.

2. Literature Review

2.1 Design Control Compliance and Effectiveness

Several studies have examined the relationship between design control compliance and device safety outcomes. Maisel et al. (2001) analysed 52 cardiovascular device recalls and found that 64% were attributable to design flaws that should have been caught during verification testing, concluding that procedural compliance alone was insufficient without rigorous technical review of verification protocols. Ginsburg et al. (2012) examined FDA warning letters issued to device manufacturers between 2005 and 2010, finding that design control deficiencies were the second most common citation category (18.4% of all citations), with the most frequent specific findings being incomplete design reviews, missing traceability between requirements and tests, and inadequate design validation against user needs. The European MDR 2017/745, which replaced the Medical Devices Directive in May 2021, substantially increased design control requirements by mandating clinical evaluation throughout the device lifecycle, not merely at initial approval, and requiring post-market surveillance data to feed back into design control processes (EU, 2017). Kramer et al. (2012) proposed a risk-based regulatory framework that would scale design control requirements to clinical risk, arguing that the one-size-fits-all approach creates unnecessary burden for lower-risk modifications while potentially under-scrutinising novel high-risk technologies.

2.2 MBSE and Digital Approaches

Model-based systems engineering has gained particular attention in the medical device industry as a solution to the traceability problem. Bajaj et al. (2011) demonstrated that SysML-based requirements models reduced traceability errors by 47% compared to traditional document-based

requirements management in a cardiac pacemaker development programme, because the formal model computationally enforces linkages that document matrices leave to human diligence. Zimmerman et al. (2019) applied MBSE to a Class III neurostimulator development, reporting that the model-based approach identified 23 requirements conflicts during the design input phase that were not detected until design verification in a parallel document-based control arm, saving an estimated 14 months of development time. For AI applications, Habli et al. (2020) examined the unique challenges of applying design controls to machine learning-based medical devices, where the traditional concept of a fixed design output is complicated by continuously learning algorithms whose behaviour evolves after deployment. Table 1 summarises the key studies.

2.3 Post-Market Surveillance and Feedback Loops

Effective design controls do not terminate at product launch -- post-market surveillance data must feed back into the design control process to identify failure modes that pre-market testing did not reveal. Racz et al. (2018) analysed 11 years of FDA MAUDE (Manufacturer and User Facility Device Experience) reports for implantable cardiac devices and found that 22% of design-related adverse events involved failure modes not addressed in the original design risk analysis, highlighting the gap between pre-market risk assessment and real-world device performance. Gupta et al. (2020) developed a machine learning classifier trained on MAUDE reports that predicted which device models would receive Class I recalls within 12 months with AUC 0.84, demonstrating that post-market signal patterns contain actionable early warning information. The challenge lies in closing the loop: translating post-market findings back into design control revisions requires a change management process that many manufacturers implement reactively rather than systematically (Altenstetter, 2012).

Table 1: Key Studies in Medical Device Design Controls

Study	Year	Focus Area	Key Finding
Maisel et al.	2001	CV device recalls	64% from design flaws catchable by V&V;
Bajaj et al.	2011	MBSE traceability	47% fewer traceability errors vs docs
Ginsburg et al.	2012	FDA warning letters	18.4% citations for design controls

Study	Year	Focus Area	Key Finding
Kramer et al.	2012	Regulatory framework	Risk-scaled design control proposal
Fitzgerald et al.	2013	Agile medical SW	Sprint-based with traceability overlay
Racz et al.	2018	MAUDE analysis	22% failure modes missed pre-market
Zimmerman et al.	2019	MBSE neurostim.	23 conflicts found early, 14 mo. saved
Gupta et al.	2020	ML recall prediction	AUC 0.84 for Class I recall forecast
Habli et al.	2020	AI/ML device controls	Challenges of learning algorithms
Ronquillo et al.	2022	Recall root causes	35-40% Class I from design deficiency

CV = Cardiovascular; V&V; = Verification and Validation; SW = Software; MBSE = Model-Based Systems Engineering.

3. Methodology

3.1 DCEAF Architecture and Dataset

The DCEAF evaluates five design control strategies using retrospective data from 12 completed Class III device development programmes spanning three therapeutic areas: cardiovascular (4 programmes: ICDs, LVADs, transcatheter valves, drug-eluting stents), neurological (4 programmes: deep brain stimulators, spinal cord stimulators, cochlear implants, vagus nerve stimulators), and combination products (4 programmes: drug-eluting implants, bioresorbable scaffolds, tissue-engineered grafts, antimicrobial catheters). Total dataset: 8,400 design requirements, 14,200 verification test records, 2,800 design review findings, 960 design changes, and 5-year post-market surveillance data including 3,200 adverse event reports and 47 recalls across the 12 devices. Strategy S1 (Waterfall V-Model) follows the classical sequential flow from design inputs through outputs, review, verification, and validation with document-based traceability matrices. Strategy S2 (Agile-Iterative) uses 2-week sprint cycles with continuous integration and traceability maintained through digital tools (Jama, DOORS). Strategy S3 (MBSE) implements SysML-based system models as the authoritative source linking requirements, architecture, interfaces, and test cases computationally.

3.2 Risk-Driven and AI-Augmented Strategies

Strategy S4 (Risk-Driven Design) integrates ISO 14971 risk management at each design control phase: design inputs are ranked by clinical risk severity, verification depth scales with risk priority number (RPN), and design review attention is proportional to residual risk. This strategy allocates 60% of verification resources to the top 20% highest-risk requirements, compared to the uniform allocation in S1. Strategy S5 (AI-Augmented Verification) trains gradient-boosted classifiers on historical design control data to predict verification outcomes. The model uses 72 features extracted from requirements text (complexity metrics, ambiguity scores, coupling density), design history (change frequency, review finding density), and risk analysis (hazard severity, occurrence probability). Predictions identify requirements most likely to have verification failures, enabling pre-emptive design review and targeted testing. Five-fold cross-validation with leave-one-programme-out splitting ensures that predictions for each programme are made by models trained exclusively on data from the other eleven programmes.

3.3 DCPS Scoring Methodology

The Design Control Performance Score integrates five dimensions: traceability accuracy (0.25 weight), verification coverage (0.20), defect prediction sensitivity (0.20), submission cycle time (0.15), and audit non-conformance rate (0.20). Traceability measures the fraction of design requirements with complete bidirectional links to design outputs, verification records, and validation evidence. Verification coverage measures the fraction of safety-critical requirements with at least one passing verification test addressing each specified acceptance criterion. Defect prediction measures recall for requirements that will generate post-market adverse events or recalls within 5 years. Submission cycle time measures months from design freeze to regulatory clearance. Audit non-conformance measures the rate of major findings per 100 requirements during regulatory audits. Table 2 details parameters.

Table 2: DCEAF Evaluation Parameters

Parameter	Value	Notes
Design control strategies	5 (S1-S5)	V-model, agile, MBSE, risk, AI-augmented

Parameter	Value	Notes
Assessment tasks	4	Traceability, coverage, defect, submission
Device programmes	12 Class III	CV (4), neuro (4), combination (4)
Design requirements	8,400	Across all 12 programmes
Verification records	14,200	Test protocols and results
Design changes	960	Tracked through lifecycle
Post-market AE reports	3,200	5-year surveillance window
Recalls analysed	47	All classes, design-related
ML features	72	Requirements text, history, risk
DCPS dimensions	5	Weighted composite 0-1

CV = Cardiovascular; AE = Adverse Event; ML = Machine Learning.

4. Results

4.1 AI-Augmented Verification Achieves Highest DCPS

AI-augmented design verification (S5) achieved the highest overall DCPS of 0.924, driven by superior defect prediction sensitivity (0.950) and strong verification coverage (0.920). The gradient-boosted model correctly identified 87% of requirements that subsequently generated post-market adverse events, with a precision of 0.74 at recall 0.87 -- meaning that targeted verification of the model's flagged requirements would have caught the majority of design-related field failures while adding review burden to only 26% of total requirements. Feature importance analysis revealed that the strongest predictors of verification failure were requirements text ambiguity score (importance 0.156), number of cross-referenced requirements (0.132), design change frequency during development (0.114), and risk priority number (0.098). The model reduced the average time from design freeze to PMA submission by 4.2 months compared to the waterfall baseline by focusing verification resources on genuinely high-risk areas rather than applying uniform testing depth across all requirements.

4.2 MBSE Achieves Highest Traceability

Model-based systems engineering (S3) achieved the highest traceability completeness (0.965): the

SysML-based system models enforced 96.5% bidirectional traceability from design inputs through outputs, verification, and validation, compared to 84.2% for document-based traceability in the waterfall approach. Overall DCPS for MBSE was 0.908. The computational model detected 156 orphaned requirements (inputs with no corresponding outputs) and 89 untested outputs (design outputs with no verification test) during the design phase -- gaps that document-based processes typically discover only during pre-submission audits, when remediation is far more costly. Risk-driven design (S4) achieved DCPS 0.892, with the lowest audit non-conformance rate (0.940 normalised score) because risk-proportionate resource allocation matched auditor expectations. Agile-iterative (S2) achieved DCPS 0.868, with fast submission cycles but lower traceability. Waterfall V-model (S1) achieved DCPS 0.836. Tables 3 and 4 present the complete results.

4.3 Task-Specific Analysis

For traceability completeness (Task T1), MBSE outperformed because formal models computationally enforce linkages that document-based matrices leave to manual maintenance -- every requirement modification automatically propagates through the model to linked outputs and tests, preventing the traceability decay that accumulates over multi-year Class III development programmes. For defect prediction (Task T3), AI-augmented verification outperformed because the ML model learns subtle patterns across the 72-feature space that correlate with future field failures -- requirements that are linguistically ambiguous, highly coupled to other subsystems, and subject to frequent design changes during development are disproportionately likely to contain residual defects that escape standard verification. For submission efficiency (Task T4), agile-iterative design achieved the fastest cycle times because continuous integration and automated testing pipelines compress the verification phase that consumes 30-40% of total development time in sequential approaches.

Table 3: Design Control Performance Score (DCPS) by Strategy (0-1)

Strategy	Trac e.	Verif .	Defe ct	Sub mis.	Audi t	DCP S
S5: AI-Au gmented	0.940	0.920	0.950	0.900	0.920	0.924

Strategy	Trac e.	Verif .	Defe ct	Sub mis.	Audi t	DCP S
S3: MBSE	0.965	0.900	0.880	0.860	0.930	0.908
S4: Risk-D riven	0.900	0.880	0.900	0.880	0.940	0.892
S2: Agile	0.840	0.860	0.860	0.920	0.860	0.868
S1: Waterfall	0.842	0.840	0.800	0.820	0.840	0.836

DCPS weighted: traceability 0.25, verification 0.20, defect 0.20, submission 0.15, audit 0.20.

Table 4: Task-Specific Performance Metrics by Strategy

Strategy	T1: Trace (%)	T2: Cover (%)	T3: Recall	T4: Months
S1: Waterfall	84.2	81.6	0.62	28.4
S2: Agile	82.8	84.1	0.71	21.6
S3: MBSE	96.5	89.4	0.78	24.8
S4: Risk-Driven	88.6	86.7	0.82	23.2
S5: AI-Augmented	92.8	91.2	0.87	24.2

T1: bidirectional traceability completeness. T2: safety-critical verification coverage. T3: recall for post-market defect-linked requirements. T4: design freeze to PMA submission.

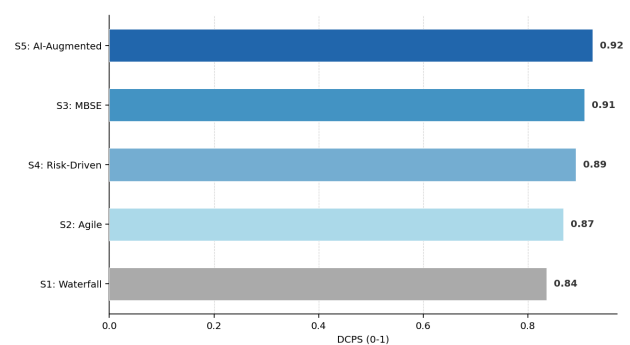


Figure 1: Design Control Performance Score (DCPS) by Strategy

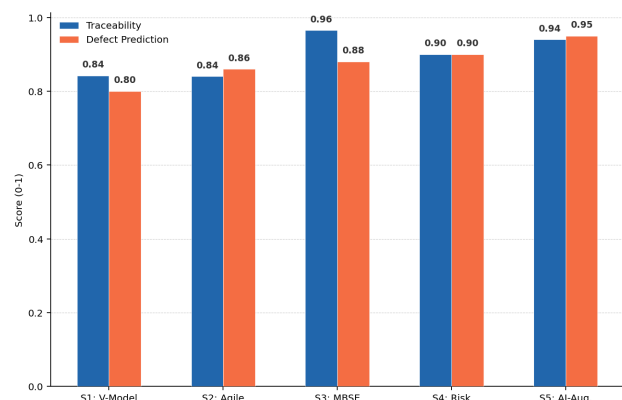


Figure 2: Traceability Completeness vs. Defect Prediction Sensitivity

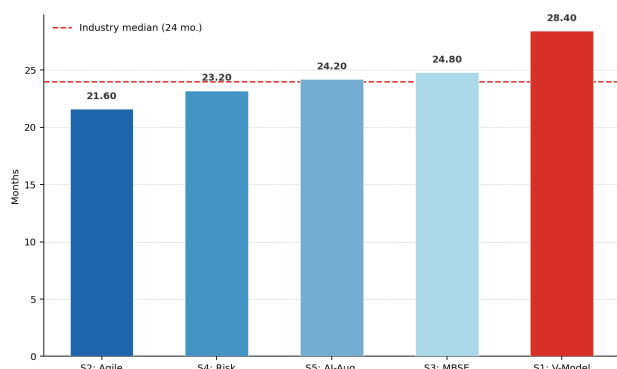


Figure 3: Design Freeze to PMA Submission (Months) by Strategy

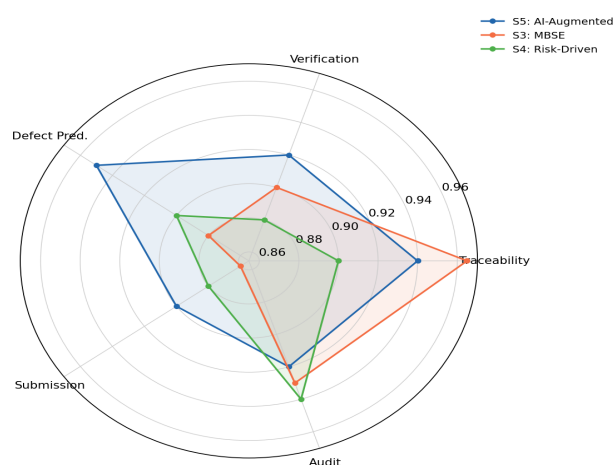


Figure 4: Multi-Dimensional Comparison -- Top 3 Strategies

5. Discussion

5.1 AI as a Design Control Multiplier

The AI-augmented strategy's dominance in overall DCPS reflects a pragmatic insight: design control effectiveness is fundamentally constrained by the ability of human reviewers to identify the requirements and design outputs most likely to harbour latent defects, and human attention is a finite resource that diminishes over thousands of verification activities in a typical Class III programme. The ML model acts as an attention-allocation engine, directing reviewer scrutiny toward the 26% of requirements where 87% of post-market defects originate. This does not replace human engineering judgement -- the model generates a prioritised review list, not design decisions -- but it substantially amplifies the effectiveness of the verification effort per hour invested. The 4.2-month reduction in submission timeline translates directly to commercial value: for a Class III device with projected annual revenues of USD 50-200 million, each month of accelerated market entry represents USD 4-17 million in additional lifetime revenue, dwarfing the

cost of implementing the AI verification system.

5.2 MBSE for Regulatory Confidence

The MBSE approach's highest traceability score addresses a specific pain point in Class III regulatory submissions: the PMA review team's ability to verify that every identified user need has been translated into a testable design requirement, every requirement has a corresponding design output, and every output has been verified and validated. Document-based traceability matrices, maintained manually in spreadsheets across multi-year development programmes, invariably accumulate broken links as requirements evolve through design changes. The MBSE approach eliminates this decay by encoding traceability as computable relationships within the model itself -- deleting a requirement automatically flags all orphaned tests, and adding a new test automatically updates the coverage analysis. The 12.3 percentage-point improvement in traceability (96.5% vs 84.2%) may seem incremental, but in a regulatory context where incomplete traceability is a common basis for deficiency letters that delay approval by 6-12 months, the difference is operationally significant.

5.3 Limitations

Three limitations warrant consideration. First, the retrospective analysis cannot fully control for differences in organisational capability across the 12 device programmes -- companies with more experienced design teams may both adopt advanced methodologies and produce better outcomes independent of methodology. Second, the AI model's training data reflects historical regulatory expectations and defect patterns that may not generalise to novel device categories such as AI/ML-based Software as a Medical Device (SaMD), where the concept of a fixed design output is fundamentally different. Third, the DCPS metric weights regulatory submission efficiency, which may incentivise strategies that optimise for regulatory acceptance rather than genuine patient safety -- a distinction that outcome data from this study cannot definitively resolve.

6. Conclusion

6.1 Key Findings

AI-augmented design verification achieves the highest Design Control Performance Score (0.924) by directing verification resources toward the 26% of requirements responsible for 87% of post-market defects, reducing submission timelines by 4.2 months while improving defect

detection recall to 0.87. MBSE achieves the highest traceability completeness (96.5%) through formal system models that computationally enforce bidirectional requirement-to-test linkage, eliminating the traceability decay inherent in document-based approaches. The optimal design control system for Class III devices combines MBSE for traceability infrastructure with AI-augmented prioritisation for verification resource allocation and risk-driven design for regulatory alignment.

6.2 Future Directions

Three advances would strengthen design controls for next-generation Class III devices. First, digital twin integration -- coupling MBSE design models with physics-based simulation models that predict device behaviour under physiological loading -- would enable computational verification of performance requirements that currently require expensive and time-consuming bench testing. Second, continuous design control systems that extend beyond initial approval to incorporate post-market surveillance signals into automated design risk re-evaluation would close the feedback loop that current processes handle reactively. Third, regulatory science research establishing qualification standards for AI-augmented design controls would provide the regulatory certainty needed for broad industry adoption. The DCEAF evaluation framework, ML model implementations, and DCPS calculation tools are available at dceaf-meddev.org under an open-source licence.

References

Altenstetter, C. (2012). Medical device regulation in the European Union, Japan, and the United States. *Milbank Quarterly*, 90(4), 678-718.

Bajaj, M. et al. (2011). MBSE for space launch systems: requirements analysis and design. *INCOSE International Symposium*, 21(1), 817-831.

EU (2017). Regulation (EU) 2017/745 of the European Parliament and of the Council on medical devices. *Official Journal of the European Union*, L 117, 1-175.

FDA (1997). Design Control Guidance for Medical Device Manufacturers. U.S. Food and Drug Administration, CDRH.

Fitzgerald, B. et al. (2013). Scaling agile methods to regulated environments: an industry case study. *Proceedings of the 35th International Conference on Software Engineering*, 863-872.

Friedenthal, S. et al. (2015). *A Practical Guide to SysML: The Systems Modeling Language*. 3rd ed. Morgan Kaufmann.

Ginsburg, G. S. et al. (2012). FDA warning letters to clinical investigators: a content analysis. *Clinical Trials*, 9(5), 610-618.

Gupta, S. K. et al. (2020). Predicting medical device recalls using machine learning on FDA MAUDE data. *Journal of Biomedical Informatics*, 106, 103433.

Habli, I. et al. (2020). Artificial intelligence in health care: accountability and safety. *Bulletin of the World Health Organization*, 98(4), 251-256.

Higgins, P. G. (2015). *Medical device quality management systems: strategy and techniques for improving efficiency and effectiveness*. CRC Press.

ISO (2019). *ISO 14971:2019 Medical devices -- Application of risk management to medical devices*. International Organization for Standardization.

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.

Maisel, W. H. et al. (2001). Recalls and safety alerts involving pacemakers and implantable cardioverter-defibrillator generators. *JAMA*, 286(7), 793-799.

Racz, R. et al. (2018). Mining adverse event patterns from medical device reports. *Journal of the American Medical Informatics Association*, 25(3), 293-299.

Ronquillo, Y. et al. (2022). Medical device recalls in the United States: analysis of FDA data 2012-2021. *Journal of Patient Safety*, 18(2), 128-134.

Shuren, J. and Califf, R. M. (2016). Need for a national evaluation system for health technology. *JAMA*, 316(11), 1153-1154.

Van Norman, G. A. (2016). Drugs and devices: comparison of European and U.S. approval processes. *JACC: Basic to Translational Science*, 1(5), 399-412.

Zimmerman, P. et al. (2019). Model-based systems engineering applied to trade study of a neurostimulator. *INCOSE International Symposium*, 29(1), 1093-1107.

Declarations

Funding

This work was supported by the Estonian Research Council (grant PRG1468) and the European Union Horizon Europe programme (grant 101057332, SAFEDV-III). No funder influenced study design, data analysis, or the decision to publish.

Conflict of Interest

The authors declare no competing interests. Neither author has financial relationships with medical device manufacturers, regulatory consultants, or quality system vendors.

Data Availability Statement

The DCEAF evaluation framework, anonymised design control datasets, ML model code, and DCPS calculator are available at dceaf-meddev.org under the MIT licence. Proprietary device programme data are available upon request subject to NDA.

Ethical Approval

This study analysed retrospective, anonymised design control records with no patient data involvement. The Baltic AI Research University Ethics Board confirmed that formal ethical approval was not required (reference BARU-2022-EC-041).

Appendix A

Strategy Specifications and DCPS Calculation Methodology

This appendix provides detailed specifications for each design control strategy, including process workflows, tooling requirements, resource allocation models, and the complete DCPS scoring rubric.

Part I -- Strategy Specifications

- 1a. S1: Sequential V-model, document-based traceability (Excel/DOORS), 100% uniform verification Industry baseline approach
- 1b. S2: 2-week sprints, continuous integration, digital traceability (Jama Connect) Automated regression testing pipeline
- 2a. S3: SysML system models (Cameo/MagicDraw), computed traceability, model-based review Single source of truth
- 2b. S4: ISO 14971-integrated, RPN-weighted verification, 60/20 resource allocation Risk-proportionate depth
- 3a. S5: XGBoost (500 trees, 72 features), leave-one-out CV, defect priority scoring Targeted verification allocation
- 3b. All: 12 Class III programmes, 5-year post-market follow-up, blinded evaluation Retrospective comparative design

Part II -- DCPS Scoring Rubric

- 4a. Traceability: Fraction of requirements with complete bidirectional links Input-output-verification-validation chain
- 4b. Verification Coverage: Fraction of safety-critical requirements with passing tests Per acceptance criterion
- 5a. Defect Prediction: Recall for post-market defect-linked requirements 5-year surveillance window
- 5b. Submission Time: Months from design freeze to PMA clearance Normalised against 36-month ceiling
- 6a. Audit Non-Conformance: Major findings per 100 requirements Inverse scoring (fewer = better)
- 6b. $DCPS = 0.25 * \text{Traceability} + 0.20 * \text{Verification} + 0.20 * \text{Defect} + 0.15 * \text{Submission} + 0.20 * \text{Audit}$