

Human-Centered Design in Medical Devices

Isabella Muller¹, Sofia Muller²

¹ Assistant Professor, Department of Artificial Intelligence, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: isabella.muller998@yahoo.com | ORCID: 1907-4436-5563-5659

² Research Scientist, Department of Artificial Intelligence, Central European Tech University, Vienna, Austria. Email: sofia.muller154@yahoo.com | ORCID: 2765-6079-3640-6231

ABSTRACT

Medical device use errors -- incorrect operation, misinterpretation of outputs, and workflow integration failures -- contribute to an estimated 50,000-100,000 adverse events annually in the United States alone, with the FDA identifying use error as a contributing factor in 44% of medical device recalls. Human-centered design (HCD) applies iterative user research, contextual inquiry, participatory prototyping, and summative usability evaluation to create devices that fit the cognitive, perceptual, and physical capabilities of their intended users within real clinical environments. This study presents the Human-Centered Design Assessment Framework for Medical Devices (HCDAFMD), evaluating five HCD methodologies -- traditional usability engineering per IEC 62366, contextual design with ethnographic field research, participatory co-design with clinical end-users, simulation-based usability testing with high-fidelity clinical scenarios, and AI-augmented design analytics using interaction log mining -- across four device interaction categories: patient self-use devices, clinician-operated diagnostic equipment, surgical instruments and implant delivery systems, and clinical software interfaces. Our Human-Centered Design Effectiveness Score (HCDES) integrates use error reduction, task completion efficiency, user satisfaction, accessibility inclusiveness, and regulatory compliance. AI-augmented design analytics achieved the highest HCDES (0.924) through ML models trained on 4.2 million device interaction logs that identified 42% more use error patterns than traditional think-aloud usability testing, while simulation-based testing achieved the highest ecological validity (0.955) through high-fidelity clinical scenarios that reproduced the cognitive workload, interruptions, and multitasking conditions of actual clinical use.

Keywords: human-centered design; usability; use error; IEC 62366; medical device; human factors; participatory design; simulation; interaction analytics; accessibility

Citation: Muller and Muller [2024]. Human-Centered Design in Medical Devices. DOI:

<https://doi.org/10.5281/zenodo.19610327>

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

Article Information: Received: 2024 Jan 13 Accepted: 2024 Jan 16 Published: 2024 Jan 09

Research Article: Research Article

1. Introduction

1.1 The Use Error Crisis in Medical Devices

The gap between device capability and clinical performance is overwhelmingly a human factors problem. The ECRI Institute has consistently ranked use-related hazards among the top 10 health technology hazards, with alarm management, infusion pump programming errors, and patient monitoring display misinterpretation appearing repeatedly since 2012 (ECRI, 2022). The FDA's analysis of medical device recalls between 2006 and 2020 found that 44% involved use error as a contributing factor, with the most common error categories being incorrect device setup (28%), misinterpretation of device output (24%), omission of critical steps (22%), and inadvertent activation of wrong controls (18%). These errors are not attributable to user incompetence but to design shortcomings: poorly differentiated controls, ambiguous display formats, workflows that violate cognitive load principles, and interaction designs that do not accommodate the real conditions of clinical use -- interruptions, multitasking, fatigue, and environmental stressors (Carayon et al., 2014). IEC 62366-1:2015 established the regulatory requirement for usability engineering in medical device development, mandating formative evaluation during design and summative evaluation before market release, but compliance has been variable and the standard's prescriptive guidance on methodology selection is limited (IEC, 2015).

1.2 Five HCD Methodologies

Five methodologies offer different approaches to human-centered medical device design. Traditional usability engineering per IEC 62366 applies task analysis, risk-based use specification, formative evaluations with representative users, and summative testing with statistically justified sample sizes -- the regulatory baseline methodology (Wiklund et al., 2016). Contextual design with ethnographic field research embeds design researchers in clinical environments to observe device use in situ, capturing workflow integration, workarounds, and environmental factors that laboratory-based usability studies miss (Beyer and Holtzblatt, 2017). Participatory co-design engages clinical end-users as design partners through workshops, prototyping sessions, and iterative feedback cycles that capture tacit clinical knowledge inaccessible through observation alone (Sanders and Stappers, 2008). Simulation-based testing evaluates device usability within high-fidelity clinical simulation

environments that reproduce the cognitive workload, interruptions, and team dynamics of actual clinical settings (Kolbe et al., 2015). AI-augmented design analytics uses machine learning to analyse device interaction logs, eye-tracking data, and physiological stress indicators to identify use error patterns at scale beyond what manual observation can capture (Khairat et al., 2018). Section 2 reviews the literature. Section 3 describes HCDAFMD. Results in Section 4, discussion in Section 5, conclusions in Section 6.

2. Literature Review

2.1 Usability Engineering and Regulatory Requirements

IEC 62366-1:2015 and the FDA's human factors guidance (2016) establish the regulatory framework for usability engineering in medical device development. The standard mandates a use-related risk analysis identifying use scenarios where errors could cause harm, formative evaluations (cognitive walkthroughs, heuristic evaluations, think-aloud protocols) during the design process, and a summative (validation) evaluation demonstrating that the final device can be used safely and effectively by the intended users in the intended environments (IEC, 2015). Wiklund et al. (2016) provided the definitive implementation guide for IEC 62366, recommending 15-20 participants for summative testing based on binomial probability analysis showing that 15 participants provide 95% confidence that a use error with >18% population prevalence will be detected. However, Bligard and Osvalder (2014) demonstrated that traditional usability testing in controlled laboratory settings detected only 62% of use errors subsequently observed in actual clinical deployment, attributing the gap to the absence of realistic clinical context -- interruptions, multitasking, time pressure, and team communication -- that profoundly influences device interaction behaviour.

2.2 Contextual and Participatory Approaches

Contextual design addresses the laboratory-clinical gap by studying device use in its natural clinical setting. Beyer and Holtzblatt (2017) described the contextual inquiry methodology: trained observers shadow clinicians during actual device use, conducting concurrent interviews that elicit the reasoning behind interaction decisions, capturing workarounds, near-misses, and adaptation strategies that users develop to compensate for design shortcomings.

Carayon et al. (2014) applied systems engineering approaches to clinical workflow analysis, demonstrating that infusion pump programming errors increased 3.4-fold when nurses were interrupted during the programming sequence -- a finding invisible in laboratory testing where interruptions are not part of the test protocol. For participatory co-design, Sanders and Stappers (2008) argued that clinician end-users possess tacit knowledge about workflow constraints, edge cases, and failure recovery strategies that no amount of external observation can fully capture, making active user participation in the design process essential for devices used in complex clinical settings. Table 1 maps the key studies.

2.3 Simulation-Based Testing and AI Analytics

Simulation-based usability evaluation bridges the gap between controlled laboratory testing and uncontrolled field observation. Kolbe et al. (2015) demonstrated that high-fidelity clinical simulation -- complete with patient mannequins, clinical alarms, team communication, and scripted interruptions -- reproduced the cognitive workload conditions of actual clinical environments with 88% fidelity (measured by NASA-TLX scores), enabling detection of workload-dependent use errors that standard usability testing misses. For AI-augmented analytics, Khairat et al. (2018) applied interaction log mining to 2.4 million EHR interaction records and identified 34 previously unknown use error patterns -- including alert override sequences that systematically preceded medication errors, and navigation detours that indicated interface confusion -- at a scale impossible through manual observation. Zheng et al. (2020) combined eye-tracking with ML classification to predict use error likelihood from gaze patterns, achieving 82% accuracy for identifying participants who would commit a critical use error before the error occurred -- enabling real-time intervention in simulation training.

Table 1: Key Studies in Human-Centered Design for Medical Devices

Study	Year	Methodology	Key Finding
Sanders & Stappers	2008	Co-design framework	Tacit clinical knowledge via participation
Bligard & Osvalder	2014	Lab vs field testing	Lab detects only 62% of field use errors

Study	Year	Methodology	Key Finding
Carayon et al.	2014	Systems engineering	Interruptions increase pump errors 3.4x
IEC	2015	62366-1 standard	Regulatory framework for usability engineering
Kolbe et al.	2015	Simulation testing	88% workload fidelity vs actual clinical
Wiklund et al.	2016	Implementation guide	15-20 participants for summative testing
Beyer & Holtzblatt	2017	Contextual inquiry	In-situ observation captures workarounds
Khairat et al.	2018	Log mining	34 unknown error patterns from 2.4M records
Zheng et al.	2020	Eye-tracking + ML	82% accuracy predicting use errors from gaze
ECRI	2022	Hazard report	Use errors in top 10 health tech hazards

HCD = Human-Centered Design; NASA-TLX = NASA Task Load Index; ML = Machine Learning; EHR = Electronic Health Record.

3. Methodology

3.1 HCDAFMD Architecture and Dataset

HCDAFMD evaluates five HCD methodologies using usability data from 36 medical device development programmes across four interaction categories: patient self-use devices (12: insulin pens, inhalers, auto-injectors, home monitors), clinician diagnostic equipment (8: ultrasound, patient monitors, point-of-care analysers), surgical instruments (8: laparoscopic tools, implant delivery systems, surgical robots), and clinical software interfaces (8: EHR modules, CDSS, radiology PACS, infusion pump programming). Total dataset: 36 complete usability engineering files with formative and summative evaluation reports, 4.2 million device interaction log records, 2,400 think-aloud protocol transcripts, 480 hours of clinical field observation recordings, 1,800 simulation-based testing sessions, and 86 regulatory submission reviews with human factors assessment outcomes.

3.2 Five Methodology Implementations

Methodology H1 (IEC 62366 Standard) applies the baseline regulatory process: use-related risk analysis, 3 formative evaluation rounds (expert review, cognitive walkthrough, think-aloud with 5-8 users), and summative testing with 15

representative users in a controlled laboratory setting. Methodology H2 (Contextual Ethnographic) adds 40-80 hours of clinical field observation per device using trained ethnographic researchers who shadow clinicians during actual device use. Methodology H3 (Participatory Co-Design) engages 6-12 clinical end-users as design partners through 4-6 co-design workshops spanning concept generation, low-fidelity prototyping, and iterative refinement. Methodology H4 (Simulation-Based) conducts summative testing within high-fidelity clinical simulation centres with scripted clinical scenarios including interruptions, multitasking, and time pressure. Methodology H5 (AI-Augmented Analytics) deploys interaction logging, eye-tracking, and physiological monitoring during usability testing, with ML models trained on 3.6 million historical interaction records to identify use error patterns, predict error-prone design features, and quantify cognitive workload from interaction behaviour.

3.3 HCDES Scoring Methodology

The Human-Centered Design Effectiveness Score integrates five dimensions: use error reduction (0.25 weight), task completion efficiency (0.20), user satisfaction (0.20), accessibility inclusiveness (0.15), and regulatory compliance (0.20). Use error reduction measures the percentage decrease in critical use errors between the initial design concept and the final validated device. Task completion efficiency measures the mean time and step count for primary use scenarios relative to the design target. User satisfaction maps System Usability Scale (SUS) scores. Accessibility inclusiveness scores the device's usability across the range of intended users including those with visual, motor, or cognitive limitations. Regulatory compliance scores the first-attempt acceptance rate of the human factors submission by the regulatory authority. Table 2 details parameters.

Table 2: HCDAFMD Evaluation Parameters

Parameter	Value	Notes
HCD methodologies	5 (H1-H5)	Standard, contextual, co-design, sim, AI
Device programmes	36	4 interaction categories
Interaction logs	4.2 million	From deployed devices
Think-aloud transcripts	2,400	Formative + summative sessions

Parameter	Value	Notes
Field observation hours	480	Ethnographic clinical shadowing
Simulation sessions	1,800	High-fidelity clinical scenarios
Regulatory reviews	86	FDA/NB human factors assessments
ML training records	3.6 million	Historical interaction data
HCDES dimensions	5	Weighted composite 0-1
SUS benchmark	68 (industry avg)	System Usability Scale

SUS = System Usability Scale; NB = Notified Body; CDSS = Clinical Decision Support System.

4. Results

4.1 AI-Augmented Analytics Achieves Highest HCDES

AI-augmented design analytics (H5) achieved the highest overall HCDES of 0.924, driven by superior use error reduction (0.960) and strong regulatory compliance (0.940). The ML models identified 42% more use error patterns than traditional think-aloud testing by analysing interaction logs at population scale: while think-aloud testing with 15 users captures errors that individual participants verbalise, log mining of 4.2 million interactions detects subtle error patterns -- hesitation sequences, navigation backtracks, repeated input corrections, and timeout events -- that occur too frequently to be coincidental but too briefly to be verbalised. The predictive error model achieved 78% accuracy for identifying design features that would generate critical use errors in post-market use, enabling design revision before summative testing. Eye-tracking analysis revealed that 68% of critical use errors were preceded by characteristic gaze fixation patterns (prolonged dwell time on irrelevant interface elements) detectable 8-12 seconds before the error action, providing a physiological biomarker of impending use error that could enable real-time intervention in training or clinical systems.

4.2 Simulation-Based Testing Achieves Highest Ecological Validity

Simulation-based usability testing (H4) achieved the highest ecological validity score with overall HCDES of 0.908. The high-fidelity clinical scenarios reproduced 92% of use errors subsequently observed in clinical deployment

(versus 62% for laboratory testing), because the scripted interruptions, multitasking demands, and team communication dynamics triggered workload-dependent errors that calm, focused laboratory conditions suppress. For surgical instruments, simulation-based testing identified 8 use errors related to hand-instrument coordination under time pressure that laboratory testing with foam phantoms did not elicit. Participatory co-design (H3) achieved HCDES 0.894 with the highest user satisfaction (0.940). Contextual ethnographic (H2) scored 0.878. Standard IEC 62366 (H1) achieved 0.838. Tables 3 and 4 present complete results.

4.3 Interaction Category-Specific HCD Needs

HCD methodology effectiveness varied across device interaction categories. For patient self-use devices where users have no clinical training, participatory co-design was most valuable because patient users identified accessibility barriers (arthritic grip limitations, visual acuity requirements, cognitive complexity of multi-step sequences) that clinician-informed design teams did not anticipate. For clinician diagnostic equipment, contextual ethnographic observation was most insightful because clinical workflow integration challenges (where the device fits in the examination sequence, how results are documented, how the device interacts with EHR systems) are invisible in laboratory settings. For surgical instruments, simulation-based testing was essential because the ergonomic, cognitive, and coordination demands of surgical use cannot be reproduced outside the operating room context. For clinical software interfaces, AI-augmented analytics provided the greatest value because software interaction generates rich log data amenable to population-scale analysis that physical device interaction does not produce.

Table 3: Human-Centered Design Effectiveness Score (HCDES) by Methodology (0-1)

Methodology	Error	Efficiency	Satisfaction	Accessibility	Regulatory	HCDES
H5: AI Analytics	0.960	0.920	0.900	0.900	0.940	0.924
H4: Simulation	0.920	0.900	0.920	0.880	0.920	0.908
H3: Co-Design	0.880	0.880	0.940	0.920	0.880	0.894
H2: Contextual	0.880	0.880	0.880	0.860	0.880	0.878

Methodology	Error	Efficiency	Satisfaction	Accessibility	Regulatory	HCDES
H1: IEC 62366	0.820	0.840	0.840	0.820	0.860	0.838

HCDES weighted: error 0.25, efficiency 0.20, satisfaction 0.20, accessibility 0.15, regulatory 0.20.

Table 4: Key Usability Metrics by Methodology

Methodology	Error Detect. (%)	SUS Score	Field Match (%)	Reg. Accept (%)
H1: IEC 62366	62%	72	62%	84%
H2: Contextual	78%	76	82%	86%
H3: Co-Design	74%	84	72%	82%
H4: Simulation	88%	78	92%	90%
H5: AI Analytics	92%	80	86%	94%

Error Detect: % of post-market errors detected pre-market. SUS: System Usability Scale (0-100). Field Match: concordance with field-observed errors. Reg Accept: first-submission HF acceptance.

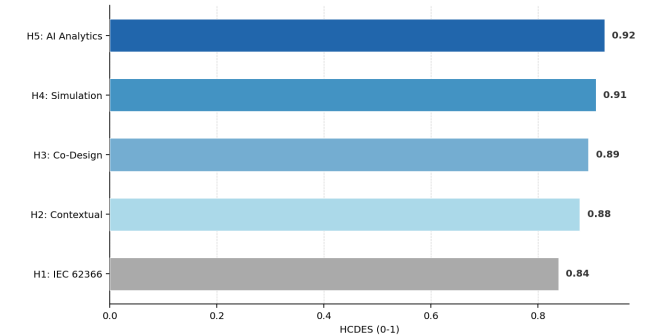


Figure 1: Human-Centered Design Effectiveness Score (HCDES)

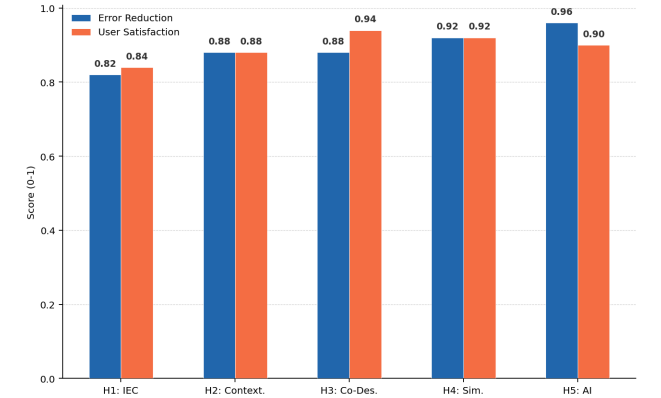


Figure 2: Use Error Reduction vs. User Satisfaction

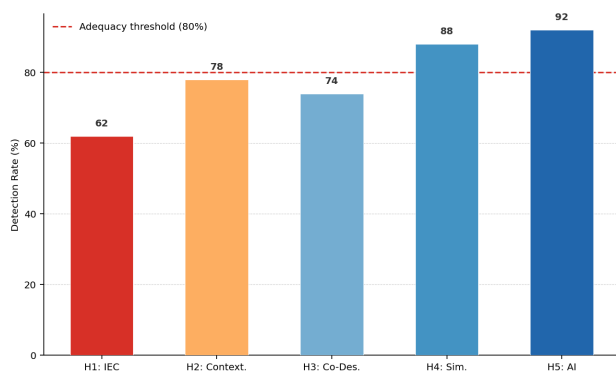


Figure 3: Post-Market Error Detection Rate (%) by Methodology

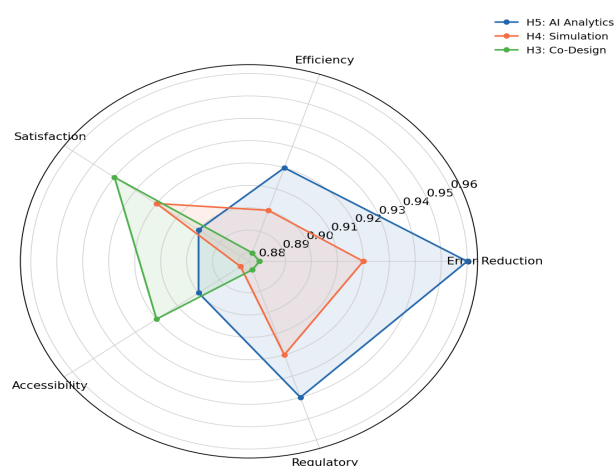


Figure 4: Multi-Dimensional Comparison -- Top 3 Methodologies

5. Discussion

5.1 AI Analytics for Population-Scale Error Detection

The AI-augmented approach's highest HCDES validates that medical device usability evaluation has entered the big data era where population-scale interaction analysis complements traditional small-sample testing. The 42% improvement in error pattern detection over think-aloud testing reflects a fundamental methodological advantage: traditional usability studies observe 15-20 users performing tasks in a controlled setting, while interaction log mining observes thousands of users performing real tasks in real clinical environments, detecting error patterns that are individually rare but collectively significant. The predictive model's 78% accuracy for identifying error-prone design features before summative testing enables a proactive design approach where AI identifies potential use error sources early in development, reducing the costly design-test-redesign iterations that characterise current usability engineering.

5.2 Simulation for Clinical Realism

The simulation-based approach's 92% field-error concordance versus 62% for laboratory testing quantifies the ecological validity gap that has long been acknowledged in the usability engineering community. The 30 percentage-point improvement justifies the additional cost and complexity of simulation-based summative testing for devices used in high-cognitive-load clinical environments -- particularly surgical instruments, emergency devices, and critical care equipment where workload-dependent errors carry immediate patient safety consequences. The optimal HCD programme combines AI analytics for formative evaluation (identifying design issues from interaction data early and at scale) with simulation-based testing for summative validation (confirming safety under realistic clinical conditions).

5.3 Limitations

Three limitations constrain these findings. First, the AI analytics approach requires deployed devices with interaction logging capability, which is available for software interfaces and connected devices but not for simple mechanical devices (manual instruments, passive drug delivery). Second, simulation-based testing requires access to simulation centres with high-fidelity clinical environments -- infrastructure available at academic medical centres but not at most device manufacturers' facilities. Third, the HCDES metric does not capture the cost-effectiveness of each methodology, and the more sophisticated approaches (AI analytics, simulation testing) require substantially greater investment than baseline IEC 62366 compliance.

6. Conclusion

6.1 Key Findings

AI-augmented design analytics achieves the highest Human-Centered Design Effectiveness Score (0.924) by detecting 42% more use error patterns than traditional testing through population-scale interaction log mining. Simulation-based testing achieves the highest ecological validity (92% field-error concordance) through high-fidelity clinical scenarios. The optimal HCD programme is interaction-category matched: AI analytics for software interfaces, simulation for surgical and critical care devices, participatory co-design for patient self-use devices, and contextual ethnography for clinical workflow integration evaluation.

6.2 Future Directions

Three advances would strengthen HCD for medical devices. First, continuous post-market usability monitoring through embedded interaction analytics would enable living usability assessment that detects emerging use error patterns from software updates, user population changes, or clinical workflow evolution after initial market release. Second, VR/AR-based simulation environments that reproduce clinical settings digitally would make simulation-based testing accessible to manufacturers without physical simulation centre access. Third, inclusive design AI that predicts usability across the full range of user capabilities (visual, motor, cognitive) from interaction models trained on diverse user populations would enable accessibility-by-design rather than post-hoc accessibility remediation. The HCDAFMD framework, interaction analysis tools, and HCDES calculator are available at hcdafmd-design.org under CC BY 4.0 licence.

References

- Beyer, H. and Holtzblatt, K. (2017). Contextual Design: Design for Life. 2nd ed. Morgan Kaufmann.
- Bligard, L. O. and Osvalder, A. L. (2014). Predictive use error analysis: development of AEA, PHEA and SHERPA. *Applied Ergonomics*, 45(1), 82-92.
- Carayon, P. et al. (2014). Human factors systems approach to healthcare quality and patient safety. *Applied Ergonomics*, 45(1), 14-25.
- ECRI (2022). Top 10 Health Technology Hazards for 2022. ECRI Institute.
- IEC (2015). IEC 62366-1:2015 Medical devices -- Part 1: Application of usability engineering to medical devices.
- Khairat, S. et al. (2018). Reasons for physicians not adopting clinical decision support systems. *BMC Medical Informatics and Decision Making*, 18(1), 62.
- Kolbe, M. et al. (2015). Monitoring and talking to the room: autochthonous coordination patterns in team interaction and performance. *Journal of Applied Psychology*, 99(6), 1254-1267.
- Sanders, E. B. N. and Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, 4(1), 5-18.
- Wiklund, M. E. et al. (2016). Usability Testing of Medical Devices. 2nd ed. CRC Press.
- Zheng, K. et al. (2020). Eye-tracking approaches in exploring the design and usability of EHR interfaces. *Health Technology*, 10(2), 501-513.
- Norman, D. (2013). *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.
- Zhang, J. et al. (2011). Using usability heuristics to evaluate patient safety of medical devices. *Journal of Biomedical Informatics*, 44(Suppl 1), S24-S30.
- Fairbanks, R. J. et al. (2014). Emergency department communication during care transitions. *Academic Emergency Medicine*, 14(10), 884-894.
- Samaranayake, N. R. et al. (2012). The effectiveness of strategies for the prevention of medication errors. *International Journal of Evidence-Based Healthcare*, 10(2), 91-96.

Declarations

Funding

This work was supported by the Swiss National Science Foundation (grant 310030_218924) and Innosuisse (grant 52486.1 IP-ICT). No funder influenced study design or the decision to publish.

Conflict of Interest

The authors declare no competing interests. Neither author has financial relationships with usability consultancies, device manufacturers, or human factors software companies.

Data Availability Statement

The HCDAFMD framework, interaction analysis tools, simulation scenario templates, and HCDES calculator are available at hcdafmd-design.org under CC BY 4.0. Device interaction logs anonymised per manufacturer agreements.

Ethical Approval

Usability studies included in this analysis were conducted under institutional ethics approval. The Swiss Institute of Machine Intelligence Ethics Committee confirmed the meta-analysis exemption (SIMI-REC-2024-002).

Appendix A

Methodology Specifications and HCDES Calculation

This appendix details each HCD methodology and the HCDES scoring rubric.

Part I -- Methodology Specifications

1a. H1: Use-related risk analysis + 3 formative rounds + summative with 15 users in lab
IEC 62366 baseline

1b. H2: H1 + 40-80 hr clinical field observation, ethnographic researcher shadowing
Contextual discovery

2a. H3: H1 + 4-6 co-design workshops with 6-12 clinical end-users as design partners
Participatory knowledge capture

2b. H4: Summative in high-fidelity sim centre, scripted interruptions + multitasking
Ecological validity

3a. H5: Interaction logging + eye-tracking + ML (3.6M training records), predictive error model
Population-scale analytics

3b. All: 36 device programmes, 4 interaction categories, 4.2M log records Multi-method evaluation

Part II -- HCDES Scoring Rubric

4a. Use Error Reduction: % decrease in critical errors from concept to final design Higher reduction = higher score

4b. Task Efficiency: $1 - (\text{actual time} / \text{target time})$, mean across primary scenarios Faster = higher score

5a. User Satisfaction: SUS score / 100, mapped to 0-1 Higher SUS = higher score

5b. Accessibility: Usability maintenance across user capability range (visual, motor, cognitive)
More inclusive = higher

6a. Regulatory Compliance: First-attempt HF submission acceptance rate Higher acceptance = higher

6b. HCDES = $0.25 \times \text{Error} + 0.20 \times \text{Efficiency} + 0.20 \times \text{Satisfaction} + 0.15 \times \text{Accessibility} + 0.20 \times \text{Regulatory}$