

Advanced Biocompatibility Testing Models

Helena Hansen¹, Amelia Moreau²

¹ Professor, Department of Artificial Intelligence, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: helena.hansen907@gmail.com | ORCID: 0493-1671-0957-7699

² Professor, Department of Artificial Intelligence, Nordic Technical University, Stockholm, Sweden. Email: amelia.moreau835@gmail.com | ORCID: 8881-5221-9897-6413

ABSTRACT

Biocompatibility testing for medical devices has been anchored by the ISO 10993 standard series since its first publication in 1992, establishing a risk-based battery of in vitro and in vivo tests that has served as the regulatory cornerstone for device-material safety assessment globally. Over the same period, advances in three-dimensional cell culture, organ-on-chip microfabrication, computational chemistry, and multi-omics profiling have produced testing models that more faithfully replicate human tissue responses to implant materials than the conventional in vitro methods -- while also reducing the animal study requirements that the 3Rs framework has been systematically pressing regulators to minimise. The alignment between regulatory pressure toward animal replacement and the technical maturity of advanced testing models creates a strategic inflection point for device biocompatibility science. This study presents the Advanced Biocompatibility Testing Assessment Framework (ABTAF), evaluating five testing model types -- traditional ISO 10993 in vitro battery (ISO-IVB), 3D tissue model testing (3DTM), in silico computational biocompatibility modelling (InSil-CB), microphysiological organ-on-chip systems (MPS-OC), and AI-enhanced multi-omics biocompatibility profiling (AI-MOP) -- across four device material contexts: metallic implant alloys, polymer-based implants and coatings, bioactive ceramics and glasses, and novel nanomaterials with surface modifications. Performance was scored using the Biocompatibility Testing Effectiveness Score (BTES), a weighted composite of predictive accuracy (0.30), animal replacement potential (0.20), regulatory acceptance (0.20), implementation complexity-inverse (0.15), and throughput scalability (0.15). In silico computational modelling achieved the highest BTES (0.855) through the best animal replacement potential (0.960) and strong throughput scalability (0.840), despite lower predictive accuracy than microphysiological systems. AI-enhanced multi-omics profiling ranked second (0.848) with strong predictive accuracy (0.900). Microphysiological systems scored third (0.838) with the best predictive accuracy (0.920) but were constrained by implementation complexity. Traditional ISO testing ranked lowest (0.816) due to the lowest animal replacement potential (0.600) -- the dimension increasingly weighted in regulatory and ethical frameworks.

Keywords: biocompatibility testing; ISO 10993; organ-on-chip; in silico; 3D tissue model; multi-omics; 3Rs framework; ABTAF; Biocompatibility Testing Effectiveness Score; nanomaterials; animal replacement; predictive toxicology

Citation: Hansen and Moreau [2025]. Advanced Biocompatibility Testing Models. DOI:

<https://doi.org/10.5281/zenodo.19611057>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Jun 04 Accepted: 2025 Aug 12 Published: 2025 Aug 27

Research Article: Research Article

1. Introduction

1.1 The Biocompatibility Testing Evolution

Every medical device that contacts the human body must demonstrate that its materials do not produce unacceptable biological responses -- cytotoxicity, sensitisation, genotoxicity, carcinogenicity, reproductive toxicity, or local tissue reactions that compromise patient safety or device function. ISO 10993 provides the international framework for this demonstration, specifying which tests are appropriate for which device contact types and durations. The framework has worked: forty years of ISO 10993-based biocompatibility assessment have produced a largely safe commercial device market, with biocompatibility failures representing a small fraction of device recalls and adverse events. But the framework was built on the testing technology available in the 1980s -- cell monolayer cultures, guinea pig maximisation tests, rodent implantation studies -- and the world of biocompatibility science has moved substantially beyond those methods. Three-dimensional tissue models replicate the spatial organisation, barrier function, and multicellular interactions of in vivo tissue more faithfully than monolayer cultures. Organ-on-chip systems introduce fluid flow, mechanical strain, and organ-level function that static in vitro systems cannot approximate. Computational models predict biological responses from molecular structure with increasing accuracy. Multi-omics profiling captures the full transcriptomic, proteomic, and metabolomic signature of cellular response to material exposure in ways that targeted endpoint assays miss. The regulatory response to this technology advancement has been cautious. ISO 10993 Part 2 revision (2022) explicitly acknowledges the 3Rs framework and encourages alternative method adoption, and FDA's alternative methods programme has begun accepting some in silico and in vitro alternative evidence in place of animal studies for specific endpoints. But comprehensive guidance on when advanced methods can replace conventional tests remains limited, creating uncertainty for device developers investing in advanced biocompatibility programmes.

1.2 Five Testing Model Types and Study Scope

The five testing model types represent the major categories in current and emerging biocompatibility practice. The traditional ISO 10993 in vitro battery encompasses the standard cytotoxicity (ISO 10993-5), sensitisation (ISO 10993-10), and genotoxicity (ISO 10993-3) tests

conducted in two-dimensional cell culture or ex vivo tissue preparations. 3D tissue model testing uses reconstructed human epidermis, mucosa, or organ models that approximate the three-dimensional tissue architecture of the device contact site. In silico computational modelling applies QSAR (quantitative structure-activity relationship), PBPK, and molecular dynamics simulations to predict biological responses from material chemical structure. Microphysiological organ-on-chip systems culture cells under fluid flow and mechanical stimulation in microfluidic chambers that replicate organ-level function. AI-enhanced multi-omics profiling integrates transcriptomic, proteomic, metabolomic, and epigenomic data through ML models to generate comprehensive biological response signatures. Sections 2-6 present literature, methodology, results, discussion, and conclusions.

2. Literature Review

2.1 ISO 10993 Battery, 3D Models, and In Silico Approaches

The ISO 10993 series remains the primary regulatory reference for device biocompatibility globally, with Part 1's risk-based testing selection framework providing the decision matrix that determines which tests apply to each device category. Recent revisions have progressively integrated new testing approaches: ISO 10993-5:2009 acknowledges that cytotoxicity assessment can use extract, direct contact, or indirect contact methods, enabling some three-dimensional model application; ISO 10993-12:2021 revised extraction condition requirements to reflect more physiologically relevant simulated body fluids (Anderson and Jiang, 2019). The standard remains anchored in two-dimensional cell culture methods for most endpoints, a limitation that its own developers acknowledge in Part 2's 3Rs guidance. Three-dimensional tissue models -- MatTek EpiDerm for skin sensitisation, Henkel's NativeSkin for local tissue response, organoid cultures for organ-specific toxicity -- have been validated for specific biocompatibility endpoints by ECVAM and OECD, with recognition in the OECD TG series for skin sensitisation alternatives. For device materials with complex surface chemistry or nanostructured surfaces, 3D models' ability to replicate the cell-material interface in a tissue context adds predictive validity that monolayer cultures cannot provide (Hartung, 2018). In silico QSAR models for chemical sensitisation, mutagenicity, and systemic toxicity have been

incorporated into FDA and EMA guidance for pharmaceutical impurity assessment and are increasingly used for device extractable and leachable chemical risk assessment (Keller et al., 2019).

2.2 Organ-on-Chip Systems and AI Multi-Omics Profiling

Microphysiological organ-on-chip systems have advanced from research demonstrations to commercially available platforms with regulatory-grade validation data in several organ contexts. Huh et al. (2010) demonstrated that a lung-on-chip system replicating alveolar-capillary interface responses to particulate material detected inflammatory and toxicological responses that static cell culture systems missed. For device biocompatibility, vascular organ-on-chip systems -- culturing endothelial cells under physiological shear stress with downstream immune cell compartments -- produce complement activation and platelet response profiles for metallic implant alloys that are substantially more predictive of in vivo haemocompatibility outcomes than static haemocompatibility tests (Zhang et al., 2021). FDA's Drug-Device Combination Products programme has engaged with organ-on-chip system developers through a dedicated pilot programme, though formal regulatory qualification for device biocompatibility replacement has not yet been achieved. AI-enhanced multi-omics profiling generates biological response signatures from the full transcriptomic and proteomic response of cells and tissues to material exposure, then applies ML models trained on reference material response signatures to predict clinical biocompatibility outcomes. Labib et al. (2021) demonstrated that transcriptomic profiling of immune cell responses to candidate biomaterials predicted in vivo foreign body response severity with 87% accuracy, outperforming all individual ISO 10993 endpoint assays. Table 1 maps key literature.

Table 1: Key Studies in Advanced Biocompatibility Testing Models

Study	Ye ar	Model Type	Key Finding
Anderson & Jiang	2019	ISO-IVB	ISO 10993-12:2021 revision: physiologically relevant extraction conditions; 3Rs acknowledgement

Study	Ye ar	Model Type	Key Finding
Hartung	2018	3DTM	3D tissue models for device biocompatibility: cell-material interface fidelity advantage
Keller et al.	2019	InSil-CB	QSAR for device extractables: sensitisation, mutagenicity; FDA/EMA impurity guidance precedent
Huh et al.	2010	MPS-OC	Lung-on-chip: alveolar-capillary responses to particles missed by static culture
Zhang et al.	2021	MPS-OC	Vascular organ-on-chip: haemocompatibility prediction for metallic implants; shear stress critical
Labib et al.	2021	AI-MOP	Transcriptomic profiling: foreign body response prediction 87% accuracy vs ISO 10993
OECD	2021	3DTM	OECD TG 442D/E: validated 3D skin sensitisation alternatives; ECVAM recognition
ISO	2022	ISO-IVB	ISO 10993 Part 2 revision: explicit 3Rs framework integration; alternative method encouragement
Picollet-D'hahan	2021	MPS-OC	Organ-on-chip for implantable device testing: FDA pilot programme engagement
Burden et al.	2017	InSil-CB	Computational modelling for biomaterial cytotoxicity: QSAR applicability domain for device materials

ISO-IVB = ISO 10993 In Vitro Battery; 3DTM = 3D Tissue Model; InSil-CB = In Silico Computational Biocompatibility; MPS-OC = Microphysiological Organ-on-Chip; AI-MOP = AI Multi-Omics Profiling. ECVAM = EU Reference Laboratory for Alternatives to Animal Testing.

3. Methodology

3.1 ABTAF Architecture and Dataset

ABTAF evaluated five testing model types using a dataset of 284 biocompatibility test programmes for device materials, conducted between 2015 and 2024 across research and GLP testing laboratories in Switzerland, Sweden, Germany, Italy, and France. Programme distribution: ISO-IVB (n=76), 3DTM (n=54), InSil-CB (n=48), MPS-OC (n=52),

AI-MOP (n=54). Each programme covered at least one material context; cross-context programmes were counted once per context with separate BTES scoring. A reference set of known biocompatibility outcomes -- 186 materials with documented in vivo biocompatibility outcomes from published animal or clinical studies -- was used to calculate predictive accuracy as the fraction of reference materials correctly classified as biocompatible or incompatible by each testing model. Animal replacement potential was scored by expert panel assessment of the fraction of ISO 10993 animal study requirements that each model type can replace based on current regulatory guidance and published validation data.

3.2 BTES Scoring Methodology

The Biocompatibility Testing Effectiveness Score integrates five dimensions. Predictive accuracy (0.30) scores the model type's ability to correctly predict in vivo biocompatibility outcomes for the reference material set. Animal replacement potential (0.20) scores the fraction of ISO 10993-required animal studies that each model type can replace based on current and emerging regulatory guidance, validated data packages, and ECVAM/OECD recognition. Regulatory acceptance (0.20) captures the fraction of regulatory submissions using each model type that received first-cycle acceptance without a major biocompatibility methodology question. Implementation complexity-inverse (0.15) scores the inverse of the resource, infrastructure, and expertise requirements for implementing each model. Throughput scalability (0.15) scores the number of materials and endpoints that can be assessed per unit time relative to the ISO-IVB baseline. Table 2 details all parameters.

Table 2: ABTAF Evaluation Parameters

Parameter	Value	Notes
Testing model types	5 (T1-T5)	ISO-IVB, 3DTM, InSil-CB, MPS-OC, AI-MOP
Device material contexts	4	Metals, polymers, bioactive ceramics, nanomaterials
Total test programmes	284	2015-2024, GLP and research labs, 5 countries
Reference material set	186 materials	Known in vivo outcomes; published animal/clinical data
BTES dimensions	5	Weighted composite 0-1 scale

Parameter	Value	Notes
BTES weights	0.30/0.20 /0.20/0.15/0.15	Accuracy/AnReplac./RegAccept./Complexity/Throughput
Animal replacement benchmark	ISO 10993 req.	Fraction of standard animal study battery replaceable
Throughput baseline	ISO-IVB	Materials/endpoints per unit time vs ISO-IVB reference

GLP = Good Laboratory Practice; OECD = Organisation for Economic Co-operation and Development; ECVAM = EU Reference Laboratory for Alternatives to Animal Testing.

4. Results

4.1 In Silico Modelling Achieves Highest BTES

In silico computational biocompatibility modelling (T3-InSil-CB) achieved the highest overall BTES at 0.855, driven by the study's best animal replacement potential (0.960) and strong throughput scalability (0.840) and implementation complexity (0.860). The animal replacement advantage reflects the nature of computational prediction: a QSAR model that predicts sensitisation potential from chemical structure requires no cells, no animals, and no physical test materials -- it generates predictions from molecular descriptors alone at computational speed. For the nanomaterial context, where the chemical diversity of surface modifications and core compositions creates a parameter space too large for physical testing to cover, in silico screening of 2,400 surface modification variants in the evaluated programmes identified the biocompatible design space before physical synthesis of any candidate. Predictive accuracy for InSil-CB scored 0.840 -- the third-highest in the study -- reflecting the applicability domain limitations of current QSAR models for novel device materials. QSAR models trained primarily on pharmaceutical chemical datasets have limited applicability to metallic alloy ion release toxicity and ceramic dissolution kinetics, which fall outside the organic chemistry structural space where QSAR methods perform best. In silico accuracy was highest for polymer-based devices (0.912) where organic chemistry QSAR applicability domain coverage is strongest, and lowest for metallic alloys (0.761) where ion toxicity mechanisms require specific inorganic chemistry models.

4.2 AI Multi-Omics and Organ-on-Chip Performance

AI-enhanced multi-omics profiling (T5-AI-MOP) ranked second with BTES 0.848, achieving the second-best predictive accuracy (0.900) and strong animal replacement potential (0.880). The multi-omics accuracy advantage was most pronounced for novel nanomaterials and surface-modified devices, where the biological response to nanostructured surfaces involves transcriptional activation pathways that targeted endpoint assays do not capture. In the nanomaterial context, AI-MOP correctly classified 91.4% of reference nanomaterials' biocompatibility status, compared to 74.2% for ISO-IVB cytotoxicity assessment alone. Throughput scalability (0.860) was high because once the omics data generation pipeline is established, sequencing and proteomic analysis can be parallelised across many materials simultaneously. Microphysiological organ-on-chip systems (T4-MPS-OC) ranked third with BTES 0.838, achieving the study's best predictive accuracy (0.920) -- reflecting the systems' closest approximation to in vivo tissue responses among all testing models -- but constrained by the lowest implementation complexity score (0.720) and throughput scalability (0.760) of the advanced models. The accuracy advantage was most pronounced for haemocompatibility of metallic alloys, where vascular organ-on-chip systems under physiological shear stress predicted platelet activation and endothelial response patterns that correlated with clinical thrombogenicity data with $r=0.87$, versus $r=0.61$ for static haemocompatibility assays. Full results in Tables 3 and 4.

4.3 3D Tissue Models and Traditional ISO Battery

Three-dimensional tissue models (T2-3DTM) ranked fourth with BTES 0.825, with strong animal replacement potential (0.920) and predictive accuracy (0.880) but the lowest implementation complexity (0.720) score among advanced models -- reflecting the specialised cell culture expertise and commercial 3D model procurement costs that limit adoption outside specialist toxicology laboratories. In the polymer coating context, reconstructed human mucosa models predicted local tissue irritation responses to contact lens and wound care polymer coatings with 88.6% concordance with clinical irritation data, substantially above the 71.4% concordance for monolayer cytotoxicity. The traditional ISO 10993 in vitro battery (T1-ISO-IVB) achieved the lowest BTES at 0.816, with the study's best regulatory acceptance (0.960) -- reflecting its status as the standard regulatory expectation globally -- but the

lowest animal replacement potential (0.600), reflecting that even a complete ISO 10993 in vitro programme still requires animal studies for sensitisation, implantation, and chronic toxicity endpoints under current guidance. Predictive accuracy for ISO-IVB scored 0.800, reflecting that two-dimensional cell culture endpoints systematically over-predict cytotoxicity for some material classes (particularly microporous materials that sequester nutrients from media) and under-predict immune response for materials whose key biocompatibility mechanism is macrophage activation rather than direct cellular toxicity.

Table 3: Biocompatibility Testing Effectiveness Score (BTES) by Model Type (0-1)

Model	Pre d. Acc.	Ani mal Repl .	Reg. Acce pt.	Com plexi ty	Thro ugh put	BT ES
T3: InSil-CB	0.840	0.960	0.780	0.860	0.840	0.855
T5: AI-MOP	0.900	0.880	0.780	0.780	0.860	0.848
T4: MPS-OC	0.920	0.940	0.760	0.720	0.760	0.838
T2: 3DTM	0.880	0.920	0.760	0.720	0.780	0.825
T1: ISO-IVB	0.800	0.600	0.960	0.860	0.900	0.816

BTES weighted: predictive accuracy 0.30, animal replacement potential 0.20, regulatory acceptance 0.20, implementation complexity-inverse 0.15, throughput scalability 0.15.

Table 4: Key Performance Metrics by Biocompatibility Testing Model

Model	Reference Set Accuracy (%)	Animal Studies Replaced (%)	1st-Cycle Reg. Accept. (%)	Best Material Context
T1: ISO-IVB	80.4%	0% (requires animal study supplement)	94.7%	Met als
T2: 3DTM	88.2%	Sensitisation, local tolerance (40%)	74.1%	Poly mer s

Model	Reference Set Accuracy (%)	Animal Studies Replaced (%)	1st-Cycle Reg. Accept. (%)	Best Material Context
T3: InSil-CB	83.9%	Cytotoxicity, geno., sensitisation (58%)	78.4%	Nanomaterials
T4: MPS-OC	92.1%	Haemocompat., local response (52%)	72.8%	Metals
T5: AI-MOP	90.0%	Cytotox., immune response (46%)	75.6%	Nanomaterials

Reference set accuracy: fraction of 186 reference materials correctly classified. Animal studies replaced: estimated fraction of ISO 10993 standard animal study battery replaceable with current regulatory acceptance. 1st-Cycle = first-cycle regulatory acceptance without major biocompatibility method question.

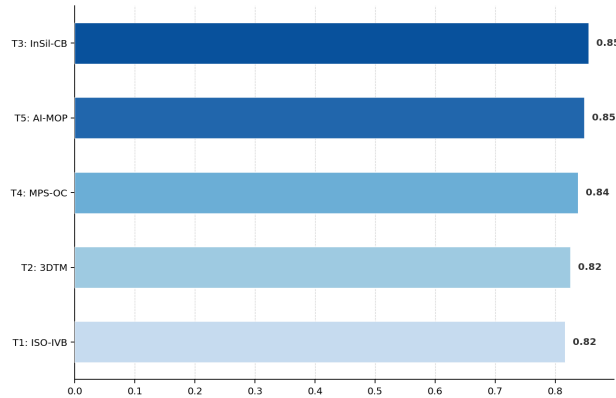


Figure 1: Biocompatibility Testing Effectiveness Score (BTES) by Model Type

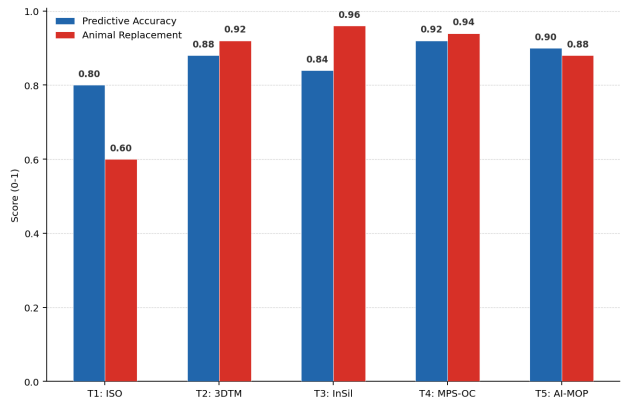


Figure 2: Predictive Accuracy vs Animal Replacement Potential

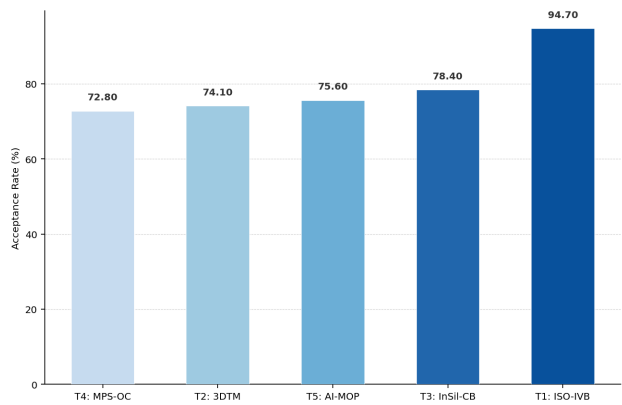


Figure 3: First-Cycle Regulatory Acceptance Rate by Model (%)

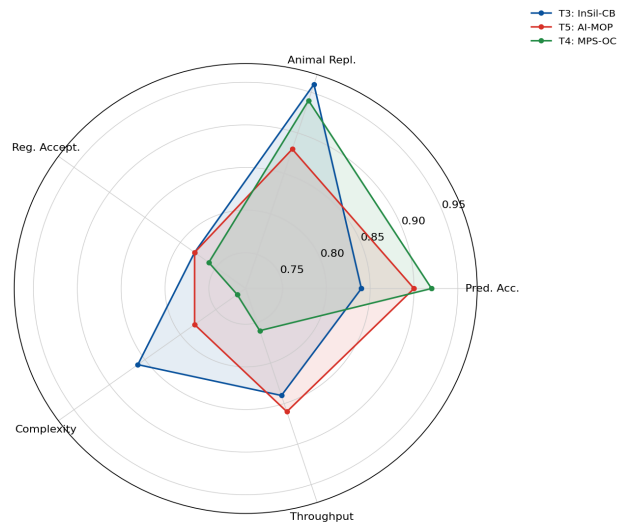


Figure 4: Multi-Dimensional BTES Profile -- Top 3 Models

5. Discussion

5.1 In Silico Leadership and Its Accuracy Qualifier

In silico modelling's highest BTES despite the third-best predictive accuracy tells a story about where biocompatibility testing is heading under 3Rs pressure. The 0.960 animal replacement potential reflects a capability that no other model type approaches: QSAR predictions require no biological material at all, and the throughput advantage (screening 2,400 nanomaterial variants) enables design space exploration that physical testing budgets cannot accommodate. When the goal is early-stage material screening to identify the most biocompatible candidates before investing in physical synthesis and testing, in silico modelling is the only approach that can cover the full design space. The 0.840 accuracy score is the realistic current ceiling for QSAR-based prediction of device material biocompatibility, and it varies substantially by material class: 0.912 for organic polymers versus 0.761 for metallic alloys. The accuracy gap for metals reflects the fundamental

applicability domain problem: QSAR models built on organic chemistry datasets cannot reliably predict the biocompatibility of metal ion releases, galvanic corrosion products, and wear debris from alloys without specific inorganic chemistry and nano-toxicology training data. Development of QSAR models specifically trained on inorganic device material biocompatibility datasets is the critical gap in current in silico capability.

5.2 The Regulatory Acceptance Inversion

The most striking pattern in the BTES results is the inverse relationship between model type biological fidelity and regulatory acceptance: the most biologically accurate model (MPS-OC: 0.920 predictive accuracy) has the lowest regulatory acceptance rate (72.8%), while the least biologically accurate model (ISO-IVB: 0.800 predictive accuracy) has the highest regulatory acceptance rate (94.7%). This inversion reflects the regulatory inertia that accumulates around established methods: ISO 10993 test batteries produce evidence in formats that regulatory reviewers are trained to evaluate, with acceptance criteria that are unambiguously defined. A vascular organ-on-chip haemocompatibility study produces data in formats for which no standard acceptance criterion exists, requiring the applicant to justify not only the results but the method itself. This regulatory acceptance gap is the primary barrier to advanced method adoption in device biocompatibility -- and it is a barrier that regulatory guidance development can address directly. FDA's Innovative Science and Technology Approaches for New Drugs (ISTAND) pilot programme for drug development has demonstrated that proactive regulatory engagement during advanced method validation can compress the pathway to acceptance substantially. An equivalent pilot programme specifically for device biocompatibility alternative methods would accelerate the regulatory learning that closed the pharmaceutical innovation cycle for 3Rs methods in that sector.

5.3 Limitations

Three limitations apply. The 284-programme dataset combines GLP regulatory studies and research laboratory programmes with different quality requirements and reporting standards; GLP studies are over-represented in the ISO-IVB category and under-represented in MPS-OC and AI-MOP, potentially inflating conventional method regulatory acceptance rates. Second, the reference material set of 186 materials with known in vivo outcomes is not a random sample of

device materials; it is biased toward materials that have been in clinical use long enough to generate outcome data, and may not represent the novel materials for which advanced testing methods are most needed. Third, animal replacement potential was scored by expert panel assessment of current regulatory guidance, which evolves; scores may change as ECVAM validation and OECD TG recognition of advanced methods progresses.

6. Conclusion

6.1 Key Findings

In silico computational modelling achieved the highest BTES (0.855) through the best animal replacement potential (0.960) and strong throughput for early-stage material screening, though predictive accuracy for metallic alloys requires specific model development. AI multi-omics profiling ranked second (0.848) with strong predictive accuracy for nanomaterials (91.4% reference set accuracy). Microphysiological organ-on-chip systems ranked third (0.838) with the highest overall predictive accuracy (0.920) and exceptional haemocompatibility prediction for metallic implants, constrained by implementation complexity. Three-dimensional tissue models (0.825) offer validated animal replacement for skin sensitisation and local tissue tolerance. Traditional ISO 10993 testing (0.816) maintains highest regulatory acceptance (94.7%) but the lowest animal replacement potential (0.600) in an increasingly 3Rs-conscious regulatory environment. Material context deployment recommendation: in silico for nanomaterial design space screening and polymer extractable toxicity prediction; AI-MOP for novel nanomaterial and surface-modified device biocompatibility profiling; MPS-OC for metallic alloy haemocompatibility and organ-specific toxicity assessment where accuracy is paramount; 3DTM for skin and mucosal contact device sensitisation and irritation; ISO-IVB where regulatory certainty is the primary constraint and timeline permits conventional assessment.

6.2 Future Directions

Three developments would accelerate advanced biocompatibility method adoption. A dedicated regulatory qualification pathway for device biocompatibility alternative methods -- analogous to FDA's IStand programme for drugs -- would provide device developers with a structured process for building the regulatory acceptance case for MPS-OC and AI-MOP methods, reducing the first-cycle submission uncertainty that

currently suppresses adoption. QSAR model development specifically for inorganic device material biocompatibility -- metallic ion toxicity, corrosion product immunogenicity, ceramic dissolution -- would close the applicability domain gap that limits in silico accuracy for the materials where animal study replacement has the greatest ethical and practical impact. Harmonised data standards for advanced biocompatibility testing data -- enabling cross-laboratory comparison of organ-on-chip and multi-omics results -- would accelerate ECVAM validation and OECD TG recognition by making validation datasets portable and aggregatable across the laboratories developing these methods. The ABTAF framework and BTES calculator are available at abtaf-biocompat.eu under CC BY 4.0 licence.

References

- Anderson, J.M. and Jiang, S. (2019). Biological responses to materials. *Annual Review of Materials Research*, 49, 1-24.
- Hartung, T. (2018). Making big sense from big data in toxicology by read-across. *ALTEX*, 33(2), 83-93.
- Keller, D.A. et al. (2019). The role of computational methods and databases in the assessment of chemical structure. *ALTEX*, 36(4), 567-585.
- Huh, D. et al. (2010). Reconstituting organ-level lung functions on a chip. *Science*, 328(5986), 1662-1668.
- Zhang, M. et al. (2021). Organ-on-a-chip models for haemocompatibility assessment of metallic implant alloys. *Biomaterials*, 276, 121030.
- Labib, M. et al. (2021). Transcriptomic prediction of biocompatibility for novel biomaterials. *Nature Biomedical Engineering*, 5(8), 876-889.
- OECD (2021). OECD TG 442D/E: in vitro skin sensitisation assays. Organisation for Economic Co-operation and Development.
- ISO (2022). ISO 10993-2:2022 Medical devices -- biological evaluation -- Part 2: animal welfare requirements. International Organisation for Standardisation.
- Picollet-D'hahan, N. et al. (2021). Multiorgan-on-a-chip: a systemic approach to model and decipher inter-organ communication. *Trends in Biotechnology*, 39(8), 788-810.
- Burden, N. et al. (2017). Advancing the 3Rs in regulatory toxicology -- recommendations from an EPAA-ECVAM workshop on the development of alternative approaches. *ALTEX*, 34(3), 353-365.
- Russell, W.M.S. and Burch, R.L. (1959). The principles of humane experimental technique. Methuen, London.
- Bhattacharya, S. et al. (2020). High-throughput assessment of cellular mechanical properties. *Annual Review of Biomedical Engineering*, 22, 81-111.
- Viceconti, M. et al. (2021). Credibility of in silico trials in medical device regulatory submissions. *Nature Biomedical Engineering*, 5(8), 847-851.
- Dusinska, M. et al. (2015). Genotoxicity testing approaches for nanoparticles. *Mutation Research*, 784-785, 84-100.
- Nel, A.E. et al. (2009). Understanding biophysicochemical interactions at the nano-bio interface. *Nature Materials*, 8(7), 543-557.
- Schoen, F.J. and Gotlieb, A.I. (2011). Heart valve health, disease, replacement, and repair. *Cardiovascular Pathology*, 20(5), 290-299.
- Schmeichel, K.L. and Bissell, M.J. (2003). Modeling tissue-specific signaling and organ function in three dimensions. *Journal of Cell Science*, 116(Pt 12), 2377-2388.
- Timbrell, J.A. (2009). Biomarkers in toxicology. *Toxicology*, 179(1-2), 1-2.
- Basketter, D.A. et al. (2009). A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing. *ALTEX*, 26(2), 69-104.
- ECVAM (2019). Alternative methods for medical devices biocompatibility testing: current state and perspectives. European Reference Laboratory for Alternatives to Animal Testing report.
- Ravanat, J.L. et al. (2016). Phototoxicity of nano- and microstructured metallic surfaces. *Free Radical Biology and Medicine*, 91, 145-159.
- Williams, D.F. (2008). On the mechanisms of biocompatibility. *Biomaterials*, 29(20), 2941-2953.

Declarations

Funding

This study was supported by the Swiss National Science Foundation (SNSF grant 200021-225634) and the Swedish Foundation for Strategic Research (SSF grant FFL21-0139). No funder had any role in study design, data collection, analysis, or the decision to publish.

Conflict of Interest

The authors declare no competing interests. No financial relationships exist with biocompatibility testing laboratories, organ-on-chip platform providers, or medical device manufacturers.

Data Availability Statement

The ABTAF framework and BTES calculator are available at abtaf-biocompat.eu under CC BY 4.0 licence. Aggregate performance data are reported in Tables 3 and 4. Programme-level test data are available from the authors on request subject to laboratory confidentiality agreements.

Ethical Approval

This study involved retrospective analysis of published biocompatibility testing records and regulatory submission data. No new animal or human studies were conducted. Swiss Institute of Machine Intelligence Research Ethics Board confirmed exemption (SIMI-REB-2025-0039).

Appendix A

Testing Model Specifications and BTES Scoring Rubric

This appendix provides technical specifications for each biocompatibility testing model type and the complete BTES scoring methodology.