

Article

Simulation as Inquiry: Reframing Methodological Knowledge Through Boundary Mapping

Sifiso Vilakati 

Department of Biostatistics, Faculty of Health Sciences, School of Biomedical Sciences, University of the Free State, Bloemfontein 9300, South Africa; sifemman@gmail.com

Abstract

Methodological statistics has developed under the pervasive belief that the value of a simulation study lies in its ability to demonstrate the superiority of a proposed method. This belief has distorted the epistemic role of simulation research by encouraging selective reporting, narrow design spaces and the suppression of results that fail to show improvement. Drawing on philosophical accounts and on meta scientific critiques by several studies, this paper argues that non-superiority is not a failure but an essential source of methodological knowledge. It marks the boundaries at which a method performs adequately, where it begins to degrade and where it breaks. Using empirical patterns evident in contemporary methodological literature and drawing on the widely cited case in which machine learning methods often failed to outperform logistic regression in clinical prediction, this paper demonstrates how non-superiority clarifies expectations and guides methodological refinement. The paper proposes a structural remedy that pairs the Registered Reports model with principled simulation design following the ADEMP framework. This alignment protects the visibility of non-superiority and restores the integrity of simulation research. Boundary mapping is therefore not a modest contribution but a scientific and ethical imperative for methodological transparency and applied reliability.

Keywords: non-superiority; boundary mapping; simulation studies; epistemic status; ADEMP framework; registered reports

1. Introduction

A longstanding assumption in methodological statistics is that scientific value is demonstrated primarily through improvement. Within this ideal of perfection, a simulation study that fails to show superiority is often viewed as a scientific disappointment rather than as an epistemic contribution. Yet the philosophical foundations of inquiry consistently reject such a narrow interpretation of progress. Popper [1] argued that knowledge advances when bold conjectures are subjected to severe testing rather than when they accumulate favourable confirmations. Kuhn and Hacking [2] described anomalies as the pressure points through which paradigms evolve and eventually break. Firestein [3] explained that ignorance and error are not liabilities but assets because they provoke the questions that drive science forward.

Contemporary metascience has amplified these classical insights. Ioannidis [4] demonstrated that in environments characterised by limited power, multiple testing and analytical flexibility, a substantial proportion of statistically significant results are likely to be false. Sterling [5] showed decades earlier that when non-significant studies remain unpublished, chance alone eventually produces significant findings that then distort the



Academic Editor: Marcin J. Schroeder

Received: 19 February 2026

Revised: 26 May 2026

Accepted: 28 May 2026

Published: 20 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

scientific archive. Loken and Gelman [6] later showed that noisy measurement exacerbates this distortion because the studies that pass significance thresholds under high noise tend to exaggerate effect sizes. Together, these arguments indicate that suppressing non-superiority corrupts the scientific record. These concerns apply acutely to simulation studies, which are uniquely positioned to test statistical tools across diverse data-generating conditions. When the literature selectively rewards superiority, simulations become demonstrations rather than explorations. This paper argues that their fundamental purpose should be boundary mapping, the identification of regions in which methods work, where they begin to deteriorate and where they fail.

This paper focuses specifically on computer-based simulation studies in statistical methodology, where methods are evaluated through repeated sampling from explicitly specified data-generating processes. While some of the philosophical arguments presented here may extend to simulation more broadly, the analysis is grounded in the role that computational simulation plays as a primary tool for method evaluation in modern statistics.

2. Locating Simulation Research in the Logic of Inquiry

Simulation occupies a distinct position within the logic of scientific inquiry. It is neither deductive mathematics, which derives analytic truths from axioms, nor empirical fieldwork, which observes naturally occurring phenomena. Instead, simulation is a logic of consequences instrument: it asks what follows from a set of statistical assumptions when those assumptions are instantiated repeatedly in controlled hypothetical worlds. This characterisation aligns with the view that knowledge advances not through accumulation of confirmations but through structured confrontation with assumptions, as emphasised in classical accounts of scientific method by Popper [1] and Kuhn and Hacking [2]. Within this framework, simulation generates conditional knowledge statements of the form: given this data-generating mechanism, these patterns of behaviour arise. The commitment to severe testing implicit in this stance, echoed in Firestein [3] on the epistemic value of failure, requires that simulations explore challenging parameter regions as part of their core logic rather than as optional embellishments.

This perspective can be sharpened by returning to Popper's central insight concerning the logical asymmetry between verification and falsification. Classical inductive reasoning seeks to confirm general laws through repeated positive instances; yet, as Popper demonstrated, such confirmation is logically inconclusive: no number of consistent observations can establish a universal claim. In contrast, falsification relies on the deductive structure of *modus tollens*: if a theory implies a particular observation, and that observation does not occur, then the theory or the specific conditions under which it is applied must be rejected or revised [1].

This logical asymmetry has profound implications for the interpretation of simulation results. In current practice, demonstrations of superiority function analogously to instances of verification; they show that a method performs well under curated conditions, but they do not establish its general validity or robustness. By contrast, instances of non-superiority play a role closer to falsification. When a proposed method fails to outperform existing approaches under well-specified conditions, the result provides a 'clash with reality' even if that reality is a simulated one, placing definitive constraints on the claims that can be made about the method. It identifies the exact boundaries where the method's assumed advantages cease to hold [1].

From this Popperian standpoint, the epistemic contribution of simulation is not exhausted by the identification of high-performing methods. Rather, its value depends fundamentally on its capacity to subject methodological claims to 'severe testing' across diverse and challenging conditions. Without systematic attention to non-superiority, simu-

lation studies risk reverting to a form of inductive confirmation, accumulating favourable cases without adequately probing the failure states. By reframing simulation as ‘Boundary Mapping,’ we shift the goal from verifying a method’s success to corroborating its resilience through the elimination of error [1].

3. The Role of Abstraction and Model Worlds in Method Evaluation

Simulation studies operate through the deliberate creation of model worlds, abstract constructs designed to isolate the structural features most relevant to methodological performance. These model worlds are not intended as replicas of specific datasets but as analytic environments that make visible the mechanisms, noise levels, dependence structures, sparsity, nonlinearity that shape estimator behaviour. This perspective is consistent with the treatment of the Data Generating Process (DGP) as the “laboratory” in which theoretical constructs are made observable, a view articulated in methodological guidance such as the ADEMP framework of Morris et al. [7]. Abstraction, in this setting, is not a simplification but a philosophical stance: it foregrounds the conditions under which performance claims hold. This accords with the boundary mapping intuition advanced throughout this paper, that methodological insight arises not merely from documenting expected strengths but from exposing the limits and breakpoints of methods under diverse, well justified abstractions.

4. Philosophical Foundations of Simulation as Inquiry

4.1. *Simulation at the Intersection of Theory, Experiment, and Instrumentation*

Computer simulation occupies a contested position within the philosophy of science. It is neither reducible to deductive theory nor fully assimilable to empirical experimentation. Instead, it operates in an intermediate epistemic space that combines elements of both while introducing distinctive problems of justification, interpretation, and trust [8]. This hybridity has generated competing philosophical accounts regarding its epistemic status.

One influential view treats simulations as extensions of theoretical reasoning. On this account, simulations simply derive the consequences of formally specified assumptions, allowing researchers to explore implications that are analytically intractable. However, this perspective underestimates the extent to which simulations rely on approximations, numerical techniques, and pragmatic adjustments that are not strictly entailed by theory. As Winsberg [9] argues, simulation practice involves layers of modelling decisions and technical interventions that possess a degree of autonomy from high-level theory. The epistemic credibility of a simulation therefore depends not only on the correctness of its assumptions but also on the reliability of its implementation and the robustness of its techniques.

An alternative view, developed by Parker [10], emphasises the experimental character of simulations. According to this interventionist perspective, what defines an experiment is not the material substrate but the act of manipulating a system to observe its response. Since simulation studies involve systematic interventions on model worlds, they can be regarded as a form of experiment conducted on constructed systems. This interpretation narrows the conceptual gap between simulation and laboratory practice and supports the idea that simulation generates genuine empirical knowledge about structured representations of the world.

Yet, as Roush [11] has argued, an important asymmetry remains. Simulations are always dependent on prior theoretical commitments because every dynamic they exhibit must be explicitly encoded. In contrast, experiments allow the natural system itself to perform causal work that exceeds the experimenter’s explicit assumptions. This structural difference implies that simulation based knowledge is conditional on the adequacy of its underlying model in a way that experimental knowledge is not.

These competing interpretations reveal that simulation cannot be understood through a single analogy. It is at once a theoretical instrument, a surrogate experimental system, and a technical artefact whose epistemic standing depends on the interaction between these roles.

4.2. Epistemic Opacity and the Problem of Understanding

A central philosophical challenge arises from the increasing complexity of modern simulations. As systems grow in scale and sophistication, they often become epistemically opaque. No individual researcher can fully survey the chain of computations that produce a given result [12]. This opacity is especially pronounced in high performance computing and machine learning systems, where methodological pathways are not only complex but dynamically evolving.

Resch and Kaminski [13] characterise this transformation as a shift from classical to trans classical technology. In classical systems, the relationship between input and output is stable and analytically accessible. In trans classical systems, the transformation rules themselves depend on training history and adaptive processes, making detailed understanding extremely difficult. The consequence is a decoupling between effectiveness and insight. A method can produce accurate predictions without offering a transparent account of how those predictions are generated.

This condition challenges traditional standards of scientific justification, which have relied on the ability to trace results back to intelligible mechanisms. When such tracing becomes infeasible, the epistemic basis of simulation shifts away from internal understanding toward external evaluation. As opacity increases, the question is no longer whether we can follow every step, but whether we have sufficient grounds to trust the process as a whole.

4.3. Trust, Entitlement, and Reliabilism

The problem of opacity has prompted a range of responses concerning epistemic trust. Symons and Alvarado [12] criticise what they describe as an entitlement view, according to which scientists may accept simulation outputs in a default manner analogous to trusting perception or memory. They argue that such an approach is inappropriate for scientific practice, where standards of justification are significantly more demanding. Treating simulations as black boxes, simply because they are successful, risks detaching scientific inquiry from its commitment to explanation and theoretical grounding.

Their critique is illustrated by the hypothetical Mechanical Oracle, a system that generates perfectly accurate predictions without offering any explanation. While such a device might be instrumentally useful, reliance on it would not constitute scientific understanding. The implication is that predictive success alone is insufficient for epistemic legitimacy.

In contrast, Durán [14] advances a reliabilist framework for simulation. He argues that justification need not depend on complete transparency but can instead be grounded in the demonstrated reliability of the computational process. On this view, simulations function as epistemic instruments analogous to microscopes or telescopes. Trust is warranted when there exists a robust chain of validation, verification, and empirical success. This includes code testing, sensitivity analysis, comparison with known results, and a history of successful application.

These positions are not mutually exclusive but highlight a tension at the heart of simulation based inquiry. On one hand, scientific norms require critical scrutiny and theoretical integration. On the other, the practical reality of complex systems necessitates forms of trust that extend beyond direct understanding. The philosophical challenge is to articulate standards that preserve rigour without demanding impossible levels of transparency [14].

4.4. Reproducibility, Instability, and Methodological Fragility

A further complication arises from the instability of complex computational systems. Even when simulations are carefully specified, small variations in implementation or numerical precision can produce divergent outcomes. In large-scale computational environments, runs conducted under ostensibly identical conditions may differ at fine levels of resolution. In machine learning, slight perturbations in input can lead to disproportionate changes in output [13].

These phenomena complicate the classical ideal of reproducibility as exact duplication. Instead, reproducibility in simulation contexts often takes the form of statistical consistency or robustness across repeated runs. This shift has important epistemic implications. It suggests that simulation results must be evaluated not as single definitive outputs but as distributions of possible behaviours across a defined space of conditions [13].

This insight aligns closely with the boundary mapping approach developed in this paper. If simulation outputs are inherently sensitive to assumptions and configurations, then their primary epistemic value lies not in isolated demonstrations but in systematic exploration of parameter spaces. The goal is not to establish that a method works in a single scenario, but to identify the regions in which it works reliably and those in which it does not.

4.5. Boundary Mapping as a Philosophical Resolution

The philosophical tensions outlined above converge on a common conclusion. Simulation based knowledge is inherently conditional, context dependent, and mediated by modelling choices [8]. It cannot be justified solely by theoretical derivation, nor solely by predictive success, nor solely by analogy with experiment. Its epistemic strength lies in structured exploration under explicit assumptions.

This perspective supports the central thesis of the present paper. Boundary mapping provides a framework that accommodates the strengths and limitations of simulation without overstating its epistemic reach. By systematically varying assumptions, documenting performance across regions of the simulation space, and making limitations explicit, simulation studies transform opacity and dependence into objects of inquiry rather than sources of hidden weakness.

In this sense, the philosophical foundations of simulation do not undermine its scientific value. Rather, they clarify its proper role. Simulation is not a replacement for theory or experiment, but a disciplined method for interrogating the consequences of assumptions. Its contribution is not the production of isolated successes but the revelation of the terrain on which methods operate.

5. Methodological Integrity as Design Discipline

Under the boundary mapping philosophy, the credibility of knowledge from simulation does not derive primarily from the numerical outcomes but from the discipline of design that produced them. This design-first perspective places epistemic responsibility on transparent declaration of aims, explicit articulation of the simulation space, careful justification of estimands, and principled selection of performance measures. Such an orientation resonates with the argument that selective visibility where only superior or favourable results are published distorts the methodological archive, a concern raised by Ioannidis [4], Sterling [5], and Loken and Gelman [6] in their analyses of publication bias and noise-driven inflation of effects. It also aligns with structural reforms that shift scientific value from outcomes to process, most notably the Registered Reports model introduced by Chambers [15] and later developed with collaborators, which binds publication decisions to methodological clarity rather than result direction. This paper adopts that ethos:

simulation is treated as epistemically sound only when its design is explicit, its assumptions defensible, and its limitations openly incorporated into the structure of inquiry.

6. Simulation Studies as the Mapping of Methodological Terrain

Simulation studies function as controlled experiments in which researchers can observe how methods behave across a variety of conditions. When conducted with transparency and rigour, they reveal the contours of methodological performance. Morris et al. [7] emphasised that principled simulation design requires pre-specification of aims, data-generating mechanisms, estimands, methods and performance measures, a structure captured in their ADEMP framework. This framework transforms simulations into scientifically disciplined experiments that resist selective reporting.

When simulations are designed and reported within this structure, the landscape becomes visible. Regions of clear superiority appear as peaks of the terrain, while areas of non-superiority appear as plateaus that indicate where established methods remain competitive. Sudden deterioration appears as cliffs where variance increases or convergence fails. Knowledge of this full landscape is essential for methodological understanding. A literature that depicts only peaks is not a map at all but a curated catalogue of successes that obscures the truth of the terrain.

From a philosophical perspective, this mapping aligns with the view that scientific knowledge grows through confrontation with the limits of theories. A simulation space that excludes difficult or unfavourable regions undermines the logic of severe testing and reduces the contribution of simulation studies to methodological understanding.

Formally, boundary mapping can be understood as the systematic exploration of a simulation design space with the explicit aim of identifying three regions:

- Regions of superiority, where a method consistently outperforms alternatives;
- Regions of non-superiority, where performance differences are negligible or absent;
- Regions of failure, where performance deteriorates or instability emerges.

This classification transforms simulation from a selective demonstration exercise into a structured exploration of methodological behaviour. Crucially, it requires that simulation designs deliberately include challenging and unfavourable conditions, rather than excluding them in pursuit of positive results.

7. The Epistemic Consequences of Suppressing Non-Superiority

The scarcity of non-superiority results in the methodological literature is a documented systemic artefact rather than a reflection of genuine methodological performance. As Sterling [5] noted decades ago, scientific publication systems selectively retain positive findings, and this logic persists in contemporary methodological statistics, where simulation studies are often expected to demonstrate improvement rather than to map methodological limits. Evidence from issues of major theoretical statistics journals consistently shows that published simulation studies overwhelmingly emphasise performance gains, with neutral or mixed outcomes appearing only rarely. These norms are not explicitly stated in editorial policies, but they emerge from a long standing cultural incentive structure: papers that fail to show superiority struggle to be framed as “contributions,” and editors and reviewers frequently interpret non-superiority as insufficient novelty rather than as valuable boundary evidence. As a result, the scientific archive disproportionately reflects peaks in the methodological landscape while obscuring the plateaus and cliffs that are essential for a complete epistemic map.

While such norms undoubtedly accelerate the refinement of high performing procedures, they simultaneously suppress the publication of essential boundary information: the plateaus where classical methods remain competitive, the cliffs where new techniques

deteriorate, and the unstable regions where estimation error or algorithmic failure emerges. Without systematic visibility into these boundaries, the literature presents an illusory terrain composed entirely of peaks, an epistemically distorted landscape that undermines severe testing, exaggerates methodological progress, and obscures the very conditions under which methods genuinely differ.

The epistemic consequences of this selective visibility are substantial. First, the suppression of non-superiority inflates the apparent performance of new methods, producing a landscape of exaggerated advances. This aligns with the finding by Ioannidis [4] that positive predictive value collapses under selectivity. Second, because noise inflates the apparent success of studies that pass significance thresholds, the literature becomes biased toward overstated claims, a process described by Loken and Gelman [6]. Third, the absence of reported failures prevents researchers from diagnosing why methods break, which parameters induce instability and which conditions limit performance.

The ethical consequences are equally serious. Applied researchers depend on methodological results when selecting tools. A literature that conceals valleys and cliffs misleads practitioners and produces decisions that are not well aligned with the limitations of methods. Methodological research therefore bears a responsibility to describe the full terrain.

8. Case Studies in Boundary Mapping

8.1. Historical Illustration: Kaplan–Meier and the Epistemic Value of Non-Superiority

The history of the Kaplan–Meier estimator provides a revealing illustration of how methodological value can be misjudged under prevailing evaluative norms. Introduced by Kaplan and Meier [16], the estimator is now foundational in survival analysis. However, its early reception was shaped by a critical perspective that prioritised efficiency relative to existing parametric methods.

Contemporary commentary characterised the estimator as incurring a statistical “price,” referring to its reduced efficiency under ideal parametric assumptions. This critique did not claim that the method failed, but rather that it did not outperform established approaches when those approaches were well specified. Implicit in this assessment is a normative expectation that new methods should demonstrate superiority as a condition of scientific contribution [17].

From the perspective advanced in this paper, this expectation is philosophically problematic. It evaluates methods primarily at points of optimal performance, while neglecting the broader range of conditions under which they may be applied. The Kaplan–Meier estimator derives its significance not from dominating parametric models in idealised settings, but from providing a robust alternative in the presence of censoring and model uncertainty. Its strength lies precisely in regions where parametric assumptions are fragile or unverifiable.

The enduring influence of the Kaplan–Meier estimator demonstrates that methodological importance cannot be reduced to efficiency comparisons under restricted conditions [18]. Rather, it emerges from the capacity of a method to extend inference into domains where existing approaches are limited. This is a form of epistemic contribution that is only visible when methods are evaluated across a sufficiently rich space of conditions.

This case highlights a broader cultural tendency within statistical methodology. The expectation that new methods must outperform existing ones remains deeply embedded in publication and review practices. Such expectations encourage the reporting of peak performance while discouraging the systematic exploration of conditions under which methods perform similarly or differently. In this way, the epistemic landscape becomes skewed toward superiority claims, at the expense of a more complete understanding of methodological behaviour.

The Kaplan–Meier example therefore supports the central claim of this paper. Had methodological evaluation been governed strictly by superiority criteria, contributions of this kind might have been undervalued or overlooked. Boundary mapping provides an alternative framework that recognises the importance of robustness, domain extension, and conditional performance, thereby aligning methodological evaluation more closely with the actual structure of scientific inquiry.

8.2. Other Examples

The comparison between logistic regression and machine learning models for clinical prediction serves as an illustrative example of a broader and recurring pattern in statistical science. For many years, it was widely assumed that machine learning methods would outperform logistic regression due to their flexibility. However, ref. [19] reviewed dozens of real-world applications and found no consistent performance benefit for machine learning when logistic regression was properly specified and validated. This result challenged a widely held methodological conjecture and clarified the conditions under which simpler models remain competitive.

This pattern is not unique to clinical prediction. Similar findings arise across multiple domains. In medical diagnostics, studies have shown that simple linear or logistic models can match or exceed the performance of more complex methods such as decision trees and support vector machines, particularly in settings characterised by small sample sizes and relatively simple predictor structures. In such cases, the flexibility of complex models introduces variance without corresponding gains in signal capture, leading to overfitting [19,20].

A parallel example is found in time-series forecasting through the Makridakis (M) competitions, where simple methods such as naïve forecasting or exponential smoothing repeatedly performed as well as, or better than, more complex autoregressive or machine learning approaches, especially for short-term predictions. These results systematically challenged the assumption that increased model sophistication guarantees improved performance [21].

These empirical patterns are underwritten by a general theoretical principle: the No Free Lunch theorem [22]. This principle implies that no single method can be universally optimal across all data-generating processes. Performance is always conditional on structure. Methods that are highly flexible gain advantage only in domains where such flexibility aligns with the underlying data-generating mechanism; otherwise, they may perform no better—or worse—than simpler alternatives.

Taken together, these examples reinforce the central claim of this paper: non-superiority is not an anomaly but a structural feature of methodological research. It reveals the conditions under which methods fail to improve upon existing approaches and therefore provides direct evidence for the boundaries of methodological claims. These boundaries are not incidental; they are the epistemic core of simulation-based inquiry.

9. The Role of Process-First Publication Structures

Transforming the simulation literature requires structural reform. The Registered Reports model introduced by Chambers [15] relocates peer review to the design stage, ensuring that publication decisions are based on the importance of the research question and the quality of the methodology rather than on the eventual results. Chambers [15] later elaborated how this format realigns publication incentives with scientific integrity by preventing authors from selectively reporting favourable results Chambers et al. [23]. Nosek and Lakens [24] argued that the Registered Reports model increases credibility by binding publication to methodological clarity.

Applied to simulation studies, Registered Reports operate synergistically with the ADEMP framework. Authors pre-commit to the simulation space they will investigate, including the regions where performance might deteriorate. Reviewers can strengthen the design before simulations are conducted, and the acceptance decision becomes independent of outcome direction. This structure legitimises the publication of non-superiority and protects the integrity of the methodological record.

Failure as Methodological Insight

Failure is often the most informative scientific event. When a simulation reveals that a new estimator offers no improvement, the result becomes a diagnostic clue. If a penalised estimator performs poorly in small samples, it suggests that penalty paths or tuning mechanisms may require revision. If a classifier overfits sparse data, the insight concerns regularisation and validation. If an estimator becomes unstable under heavy tailed distributions, the opportunity arises to explore robust loss functions. Each of these improvements begins with the honest documentation of non-superiority.

The neutral comparison tradition has long argued for this orientation, insisting that scientific progress depends on transparency in performance evaluation. When the literature embraces non-superiority, it embraces the raw material of methodological refinement.

10. Conclusions

Simulation studies are central to methodological science. They serve not merely as demonstrations of performance but as instruments for mapping the conditions under which methods succeed, under which they weaken and under which they fail. A literature that records only peaks presents a distorted scientific landscape. The suppression of non-superiority produces epistemic exaggeration, impedes methodological refinement, and undermines applied decision making.

A process-first approach that integrates Registered Reports with principled simulation design restores the integrity of simulation studies. It shifts the focus from outcome to design and permits the honest publication of non-superiority. Boundary mapping is thereby recognised as a scientific and ethical imperative. In a community committed to truth and accountability, non-superiority is not an embarrassment. It is a contribution.

It is important to emphasise that this argument does not reject the pursuit of methodological improvement or the value of demonstrable success. High-performing methods are essential for scientific and applied progress. However, a literature that records only successes provides an incomplete and potentially misleading account of that progress. Boundary mapping complements, rather than replaces, the search for superior methods by ensuring that the conditions under which success is achieved are properly understood. The concept of boundary mapping emerges naturally once simulation is understood as a method constrained by assumptions, shaped by epistemic opacity, and justified through disciplined design rather than outcome alone.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. This manuscript presents a theoretical framework and does not report empirical data.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Popper, K. *The Logic of Scientific Discovery*; Routledge: Boca Raton, FL, USA, 2005.
2. Kuhn, T.S.; Hacking, I. *The Structure of Scientific Revolutions*; University of Chicago Press: Chicago, IL, USA, 1970; Volume 2.
3. Firestein, S. *Failure: Why Science Is so Successful*; Oxford University Press: Oxford, UK, 2015.
4. Ioannidis, J.P. Why most published research findings are false. *PLoS Med.* **2005**, *2*, e124. [[CrossRef](#)] [[PubMed](#)]
5. Sterling, T.D. Publication decisions and their possible effects on inferences drawn from tests of significance—Or vice versa. *J. Am. Stat. Assoc.* **1959**, *54*, 30–34. [[CrossRef](#)]
6. Loken, E.; Gelman, A. Measurement error and the replication crisis. *Science* **2017**, *355*, 584–585. [[CrossRef](#)] [[PubMed](#)]
7. Morris, T.P.; White, I.R.; Crowther, M.J. Using simulation studies to evaluate statistical methods. *Stat. Med.* **2019**, *38*, 2074–2102. [[CrossRef](#)] [[PubMed](#)]
8. Grüne-Yanoff, T.; Weirich, P. The philosophy and epistemology of simulation: A review. *Simul. Gaming* **2010**, *41*, 20–50. [[CrossRef](#)]
9. Winsberg, E. *Science in the Age of Computer Simulation*; University of Chicago Press: Chicago, IL, USA, 2019.
10. Parker, W.S. Does matter really matter? Computer simulations, experiments, and materiality. *Synthese* **2009**, *169*, 483–496. [[CrossRef](#)]
11. Roush, S. The epistemic superiority of experiment to simulation. *Synthese* **2018**, *195*, 4883–4906. [[CrossRef](#)]
12. Symons, J.; Alvarado, R. Epistemic entitlements and the practice of computer simulation. *Minds Mach.* **2019**, *29*, 37–60. [[CrossRef](#)]
13. Resch, M.; Kaminski, A. The Epistemic Importance of Technology in Computer Simulation and Machine Learning. *Minds Mach.* **2019**, *29*, 9. [[CrossRef](#)]
14. Durán, J.M. *Computer Simulations in Science and Engineering: Concepts-Practices-Perspectives*; Springer: Berlin/Heidelberg, Germany, 2018.
15. Chambers, C.D. Registered reports: A new publishing initiative at Cortex. *Cortex* **2013**, *49*, 609–610. [[CrossRef](#)] [[PubMed](#)]
16. Kaplan, E.L.; Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481. [[CrossRef](#)]
17. Miller, R.G., Jr. What price kaplan-meier? *Biometrics* **1983**, *39*, 1077–1081. [[CrossRef](#)] [[PubMed](#)]
18. Ryan, T.P.; Woodall, W.H. The most-cited statistical papers. *J. Appl. Stat.* **2005**, *32*, 461–474. [[CrossRef](#)]
19. Christodoulou, E.; Ma, J.; Collins, G.S.; Steyerberg, E.W.; Verbakel, J.Y.; Van Calster, B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J. Clin. Epidemiol.* **2019**, *110*, 12–22. [[CrossRef](#)] [[PubMed](#)]
20. Lim, T.S.; Loh, W.Y.; Shih, Y.S. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Mach. Learn.* **2000**, *40*, 203–228. [[CrossRef](#)]
21. Makridakis, S.; Hibon, M. The M3-Competition: Results, conclusions and implications. *Int. J. Forecast.* **2000**, *16*, 451–476. [[CrossRef](#)]
22. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **2002**, *1*, 67–82. [[CrossRef](#)]
23. Chambers, C.D.; Dienes, Z.; McIntosh, R.D.; Rotshtein, P.; Willmes, K. Registered reports: Realigning incentives in scientific publishing. *Cortex* **2015**, *66*, A1–A2. [[CrossRef](#)] [[PubMed](#)]
24. Nosek, B.A.; Lakens, D. A method to increase the credibility of published results. *Soc. Psychol.* **2014**, *45*, 137–141. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.