







# The case for embracing trial-and-error in health research: Rethinking failure and uncertainty as foundations of progress

Stefan Gaillard <sup>1,2</sup>, Stefan van Geelen <sup>3</sup>, Elvire Landstra <sup>2,4</sup>, Arno Hoes <sup>5</sup>

**Keywords** *failure, uncertainty, health domain, trial-and-error*

Uncertainty is a daily reality in the health domain. Clinicians routinely face it when diagnosing patients, predicting outcomes, and selecting treatments in partnership with patients. Patients and their families, in turn, experience uncertainty about their condition and what the future may hold. Biomedical and clinical researchers aim to reduce this uncertainty through rigorous research, but not all of it succeeds. As others have argued, tolerating uncertainty may be one of medicine's next revolutions (Simpkin & Schwartzstein, 2016), yet the process of trial-and-error in science remains underacknowledged and underexplored. Sharing failures can save time and resources, ultimately improving patient outcomes. Still, most failed research disappears into the proverbial file drawer (Feng et al., 2024; Turner et al., 2022), leaving valuable lessons unlearned.

The international movement to “increase value and reduce waste” in health research has highlighted systemic inefficiencies at every stage of the research cycle – from question formulation to publication and implementation (Macleod et al., 2014). At the same time, translational medicine continues to face what is often called the “valley of death”: the persistent gap between promising biomedical discoveries and their successful application in clinical practice (Seyhan, 2019). What emerges from the contributions in this special issue is that failure

in health research is not always due to simple incompetence or lack of rigor, but rather often the result of complex, context-dependent dynamics: interventions that worked elsewhere fall flat in new cultural or institutional settings; sound experimental designs encounter logistical or ethical barriers; promising innovations reveal unanticipated side effects, diminishing returns, or practical incompatibilities when scaled. We therefore believe that one under-recognized source of waste is the neglect of *rigorous failure*: research that is methodologically and logistically well-designed and well-justified, but does not lead to successful outcomes. This includes, on the one hand, studies that yield null or negative results — such as the absence of an expected effect — and, on the other hand, well-reasoned methodological or logistical innovations that ultimately prove unworkable in practice. Both types of failure can offer crucial insights for researchers facing similar uncertainties yet are rarely shared or systematically analyzed. This opacity surrounding failure is driven not only by cultural discomfort, but also by institutional disincentives — funding structures which leave no room to publish failure, journals which reject failed research more often than not, and tenure tracks which penalize failure.

A pattern across the studies published in this issue is that failure often emerges at the

<sup>1</sup>Institute for Science in Society, Radboud University Nijmegen, the Netherlands

<sup>2</sup>Center of Trial and Error, Utrecht, the Netherlands

<sup>3</sup>Education Center, University Medical Center, Utrecht, Netherlands

<sup>4</sup>Center of Research on Psychological disorders and Somatic diseases (CORPS), Department of Medical and Clinical Psychology, Tilburg University, The Netherlands

<sup>5</sup>University Medical Center Utrecht, Utrecht, the Netherlands

## Part of Special Issue

Scientific Failure and Uncertainty in the Health Domain

## Received

July 2, 2025

## Accepted

July 9, 2025

## Published

September 14, 2025

## Issued

May 24, 2024

## Correspondence

gaillard@trialanderror.org

## License

This article is licensed under the **Creative Commons Attribution 4.0 (CC-BY 4.0)** license, which allows you to copy, redistribute, transform and build upon the article and all its contents in any medium or format for any purpose, provided that appropriate credit is given.

© Gaillard et al. 2025



point of translation: where evidence, ideas, or technologies must move between different domains — between lab and clinic, between cultures, between trial protocol and real-world practice. This translational friction reveals that success in one context does not guarantee generalizability. Another recurring insight is that *failure often exposes implicit assumptions* — about how people behave, how systems function, or how knowledge travels — that only become visible when things do not work as expected. These productive failures are especially instructive, as they force researchers to recalibrate their models, expectations, and designs. Importantly, these productive failures are often generalizable to other researchers attempting similar translations. By reflecting on both small setbacks and larger failures, we aim to foster a culture of continuous improvement that enhances scientific progress throughout the research cycle. Through the contributions in this issue, we highlight how embracing failure as a constructive and integral part of research can lead to more innovative solutions and thus better patient outcomes and improved health.

Publishing failed studies offers several key advantages that can significantly improve the efficiency and quality of health research. One of the most direct benefits is the prevention of duplicate work. When failed studies are shared, researchers can learn from one another, ensuring that they do not repeat the same unsuccessful lines of inquiry. This not only saves time but also prevents the waste of valuable funding. Avoiding redundant efforts is thus crucial for advancing patient care.

Another important advantage is the ability to gain insights into how things *do* work by learning from the failures of others. Often, failures contain valuable lessons that other studies may overlook. By understanding why certain approaches did not work, researchers can refine their hypotheses, improve their study designs, and ultimately avoid pitfalls. This process of learning from failure accelerates innovation and fosters a deeper understanding of complex systems, especially in scientific disciplines where the path to success is not always straightforward.

Moreover, sharing negative or null results contributes to more robust meta-analyses. Scientific knowledge is rarely derived from a single

study; it typically emerges from the accumulation and synthesis of data across multiple, rigorously conducted studies. When negative or null results from rigorous studies are made available, they become part of this broader analysis, offering a more complete picture of the evidence. Meta-analyses that include all types of results provide a clearer, more reliable view of a research question, allowing for more accurate conclusions and better-informed decisions in healthcare.

Finally, the increasing employment of machine learning in research benefits from the publication of failed studies. Machine learning algorithms rely heavily on large, diverse datasets to train and optimize models. When failures are included in these datasets, they enhance the quality of the machine learning output by providing the necessary variety of outcomes that can lead to more accurate predictions and solutions. By incorporating both successes and failures, machine learning models can become more robust and balanced, and thus more useful for research and clinical applications.

## I Institutional context

This special issue is a collaborative effort between the Journal of Trial and Error and *The New Utrecht School*, sponsored by the Social Innovation Program at the University Medical Center (UMC) Utrecht. While the conceptualization of this issue took place in Utrecht, the Netherlands, contributions span a broad range of research institutions from multiple countries, reflecting the global relevance of failure and uncertainty in health research. These contributions highlight the diverse ways in which failure and uncertainty manifest across various health disciplines, settings, and research contexts. The contextual nature of failure and uncertainty means that their impact can vary greatly depending on the specific circumstances, making it essential to reflect on these variations. For this reason, we have paired empirical articles with reflection pieces to foster a deeper understanding of how failure can be approached and learned from across different contexts.

Parallel to the publication process, we organized a series of monthly lunch lectures focused on scientific failure and uncertainty

in the health domain. These lectures served as an important complement to the Adrienne Cullen Lecture, which already exists at the UMC Utrecht as a prominent forum for discussing failures in clinical practice — but not for discussing failure in biomedical research. The lunch lectures provided an informal yet intellectual space for researchers, clinicians, and students to engage in discussions about how failure and uncertainty are addressed in health research and practice. These conversations fostered a deeper appreciation of the role of failure in the scientific process and encouraged participants to reflect on the potential benefits of embracing failure on research outcomes, clinical practices, and health outcomes.

### Contributions

*How to fail successfully* by Berent Prakken offers a personal and reflective account of failure in translational medicine, centered on a collapsed biotech collaboration around peptide immunotherapy for arthritis. Through vivid narrative and contextual insight, Prakken argues that failure should be understood as an instance of knowledge creation. Drawing from clinical, entrepreneurial, and educational experiences, he advocates for training translational scientists to embrace uncertainty, complexity, and setbacks as part of their professional identity and ethical responsibility.

*Issues in clinical studies leading to Medical Research Ethics Committee (MREC) negative decisions* by Heinsbroek et al. investigates the reasons behind negative decisions issued by MREC NedMec on clinical research proposals over a 5-year span. The study identifies frequent shortcomings such as incomplete research files, weak or missing scientific rationale, inadequate benefit-risk analyses, and insufficient or unclear consent procedures. These issues often lead to rejections, but many of these problems are preventable by submitting well-prepared, compliant proposals. The authors recommend that researchers consult with ethicists, methodologists, or regulatory bodies early in the design process to improve the likelihood of approval and to uphold ethical standards.

In *The “function” of art?: Challenges of setting up artistic research residencies in elderly care institutions*, Hübner et al. examine three

6-month artistic residencies in a Dutch elderly care home, where artists explored how art could address relational shifts during residents’ transition to institutional care. The project faced challenges, including staff expectations, reliance on language over art, and lack of creative space. Despite these tensions, the authors argue that art’s “functionless” nature allows for open-ended engagement, shifting how healthcare institutions view meaning, connection, and care. They call for more thoughtful project design, earlier engagement with art, and dedicated spaces for creative practices in social care settings.

A large, preregistered longitudinal experiment conducted by Panizza et al. during the 2021 COVID-19 UK vaccination campaign, *Medical expert endorsement fails to reduce vaccine hesitancy in U.K. residents*, contradicts findings from a similar Italian study. Participants received messages correcting vaccine misconceptions, each endorsed by medical experts. Contrary to prior findings from a similar Italian study, the expert-backed messages had no significant effect on participants’ vaccination intentions, beliefs, or uptake. The authors discuss possible explanations, including timing of the UK campaign, cultural context, and ceiling effects due to already high baseline vaccine support. They conclude that expert endorsement alone may not be a universally effective strategy and emphasize the need for more context-sensitive interventions.

In *On the significance of place: Vaccination refusal as a situated phenomenon*, Martijn van der Meer continues and deepens this reflection. Interpreting the UK trial as a conceptual replication, Van der Meer argues that expert messaging is not universally effective and critiques the “knowledge deficit model” that assumes hesitancy stems from ignorance. Drawing from medical humanities and anthropology, he proposes that trust, local context, and social dynamics shape how vaccine refusal operates. Public health campaigns, he concludes, should adopt more dialogical, participatory strategies tailored to specific communities instead of assuming one-size-fits-all scientific messaging will succeed.

David Grüning’s article *Digital nudges: A reflection on challenges and improvements inspired by the Gloria Adherence Subproject* offers a critical reflection on digital nudging in-

interventions, using the Gloria Adherence Subproject as a case study. The original project, reported in an earlier issue of the Journal of Trial and Error, aimed to improve medication adherence via a daily app reminder yet showed no significant effect. Grüning identifies three broader challenges of digital nudges: they can undermine user agency, their effects decay quickly over time, and they often fail in complex (“wicked”) environments. He proposes “boosting” as a more sustainable alternative — interventions that build user competencies. Grüning recommends shifting from simple reminders to empowering participants with clear information and adaptive tools.

Tackling a popular “biohacking” idea, Batorek and Hromatko examine the cognitive impacts of a 2-month time-restricted eating (TRE) regimen in *Cognitive functions, mood and sleep quality after two months of intermittent fasting*. The study tested the effects of TRE on cognitive performance, mood, and sleep in healthy adults. Although the experimental group lost weight, no statistically significant differences were found between the fasting and control groups on a battery of cognitive tasks or self-reported measures of mood and sleep quality. Improvements observed across both groups were attributed to practice effects or seasonal mood changes. The authors urge caution against overhyping the cognitive and psychological benefits of intermittent fasting and highlight the need for better-controlled, randomized studies before endorsing TRE as a cognitive or mental health intervention.

In his commentary *Cognitive or emotional improvement through intermittent fasting? Reflections on hype and reality*, Stephan Schleim reviews the previous study, highlighting methodological limitations such as the short intervention duration, lack of randomization, and limited ecological validity of cognitive tests. He contextualizes the findings within broader debates on neuroenhancement and cautions against inflated expectations, noting that even pharmacological interventions show only modest cognitive effects in healthy individuals. Schleim concludes that while intermittent fasting remains a topic of public and scientific interest, reliable evidence for its mental benefits remains lacking and nuanced, long-term research is needed.

The exploratory pilot study *Smile, you’re on*

*camera: Investigating the relationship between selfie smiles and distress* by Lind et al. investigated whether smiling behavior captured in selfie videos via a mobile sensing app (EARS) reflected self-reported psychological distress in college students. Contrary to the hypothesis based on Paul Ekman’s expressed emotion framework — that smiling decreases as distress increases — the study found that smiling intensity either stayed the same or even increased with higher levels of stress, anxiety, and depression. These counterintuitive results align more with Alan Fridlund’s behavioral ecology view, which posits that facial expressions serve communicative, rather than expressive, functions. The study highlights limitations such as small sample size, lack of positive affect measures, and simplistic automated smile detection, but nonetheless contributes to debates about the validity of facial expressions as emotional indicators. The findings suggest that smiling may serve social or self-regulatory functions under stress and demonstrate the potential and the pitfalls of mobile sensing and automated facial analysis in psychological research.

Reitsema et al. provide a reflective commentary on this work, interpreting its counterintuitive finding through the lens of constructed emotion in *A smiling paradox: Exploring the constructed nature of emotions. A reflection on the relationship between smiling in selfies and distress*. Challenging Paul Ekman’s idea of universal, biologically fixed emotional expressions, the authors align with Lisa Feldman Barrett’s view that emotions are context-dependent and shaped by past experiences, social norms, and individual goals. They argue that the act of recording oneself likely heightens self-awareness and social pressure to appear positive, prompting smiles that do not reflect genuine emotional states. The commentary also critiques the oversimplified assumptions behind AI-powered emotion recognition systems, warning against the ethical and practical risks of interpreting facial expressions as direct indicators of emotion. Ultimately, Reitsema et al. call for more nuanced, context-sensitive methods in both research and technology to capture the true complexity of human emotional life.

In *Prenatal sildenafil and fetal-placental programming in human pregnancies complicated by fetal growth restriction: A retrospective gene*

*expression analysis*, Terstappen et al. describe their study which investigated the impact of prenatal sildenafil treatment on placental and fetal gene expression in pregnancies affected by fetal growth restriction (FGR). Despite the STRIDER trial showing no clinical benefit from sildenafil in FGR, this study reveals that sildenafil does exert biological effects on gene expression relevant to fetal development and long-term cardiovascular and renal programming. Specifically, the authors used gene expression and enrichment analyses to identify changes in pathways related to smooth muscle proliferation, DNA repair, vascular development, and kidney function.

The accompanying reflection focuses on the value of studies that capture the often subtle and complex molecular effects that can occur even when no gross differences are observed. In their piece *In the era of whole transcriptome sequencing: Reflections on the molecular genetic effect of prenatal sildenafil for fetal growth restriction*, Bakhuis and Van der Heyden emphasize these effects on placental and fetal gene expression. Their reflection highlights the importance of such mechanistic studies, recommends broader analytic approaches (e.g., epigenetics, proteomics), and advocates for integrating molecular profiling into all clinical trials — especially when repurposing drugs — to better understand their multifaceted effects and to inform safer, more effective treatments for FGR.

In *Partial Endothelial Trepanation versus Deep Anterior Lamellar Keratoplasty in keratoconus patients: Results of the PENTACON trial*, Wisse et al. document the trial-and-error process of the PENTACON trial, a multicenter randomized study comparing Deep Anterior Lamellar Keratoplasty (DALK) with the newer Partial Endothelial Trepanation (PET) technique in keratoconus patients. The trial was prematurely terminated due to under-enrollment, evolving clinical practices, and loss of equipoise. Despite technical promise, both techniques showed unexpectedly high complication rates and a steep learning curve, with no clear safety advantage for PET over DALK. Although underpowered, the study illustrates critical challenges in surgical trials — balancing methodological rigor, innovation, and patient safety — and highlights the ethical responsibility to

report inconclusive or halted trials to advance evidence-based practice.

In *Reflection on the PENTACON Trial: Lessons learned from an unpublished study*, Robert Wisse candidly reflects on the failure of the PENTACON clinical trial, which was terminated due to poor recruitment, protocol rigidity, and loss of clinical equipoise. Beyond the halted trial, Wisse draws valuable lessons about trial design, feasibility, mentorship, and the ethical imperative to publish negative results. His experience underscores the need for adaptive protocols, stronger support for early-career researchers, and structural reforms to mitigate publication bias. While scientifically inconclusive, the PENTACON trial became a powerful teacher, shaping Wisse's more resilient and ethically driven research ethos.

## Conclusion

Although the focus of this special issue is on health science, the taboo surrounding failure is a broader cultural problem. Changing the culture surrounding failure will be a difficult endeavor, as cultural change often is. One crucial component for cultural change is education. By fostering an environment in which failure is recognized not as a setback but as an opportunity for learning, we can reshape how both researchers and the wider public perceive scientific progress. Encouraging open dialogue about failures, incorporating failure as a learning tool in academic curricula, and supporting transparency in research will be essential steps in shifting this cultural norm.

Ultimately, embracing failure in science can catalyze more robust, innovative solutions and improve the translation of research into real-world benefits. As the studies in this issue demonstrate, rigorous failures, when shared and understood, offer valuable insights that can accelerate progress and improve patient outcomes. The path forward will require effort, honesty and commitment, but the increased value — a greater transparency, more effective research, and better-informed health decisions and outcomes — are well worth the challenge.

## References

Feng, Q., Mol, B. W., Ioannidis, J. P., & Li, W.



(2024). Statistical significance and publication reporting bias in abstracts of reproductive medicine studies. *Human Reproduction*, 39(3), 548-558. <http://doi.org/10.1093/humrep/dead248>

Macleod, M. R., Michie, S., Roberts, I., Dirnagl, U., Chalmers, I., Ioannidis, J. P., Al-Shahi Salman, R., Chan, A., & Glasziou, P. (2014). Biomedical research: Increasing value, reducing waste. *The Lancet*, 383(9912), 101-104. [http://doi.org/10.1016/S0140-6736\(13\)62329-6](http://doi.org/10.1016/S0140-6736(13)62329-6)

Seyhan, A. A. (2019). Lost in translation: The valley of death across preclinical and clinical divide—identification of problems and overcoming obstacles. *Translational Medicine Communications*, 4(1), 1-19. <http://doi.org/10.1186/s41231-019-0050-7>

Simpkin, A., & Schwartzstein, R. (2016). Tolerating uncertainty — the next medical revolution? *New England Journal of Medicine*, 375(18), 1713-1715. <http://doi.org/10.1056/NEJMp1606402>

Turner, E. H., Cipriani, A., Furukawa, T. A., Salanti, G., & de Vries, Y. A. (2022). Selective publication of antidepressant trials and its influence on apparent efficacy: Updated comparisons and meta-analyses of newer versus older trials. *PLOS Medicine*, 19(1), Article e1003886. <http://doi.org/10.1371/journal.pmed.1003886>