ORIGINAL RESEARCH PAPER
Volume - 15 | Issue - 05 | May - 2026 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

THE LAST BENCHMARK: THE NEED FOR A NEW EVIDENTIARY PARADIGM

Pharmacology

Jiya Doshi

ABSTRACT

In 1993, five patients died and two others required emergency liver transplants during a Phase II clinical trial of fialuridine, a nucleoside analogue developed to treat hepatitis B. This observed toxicity was not predicted by preclinical studies in laboratory animals with no evidence of hepatic, pancreatic, skeletal-muscle, or nerve damage. This is because standard animal models cannot replicate the processing of nucleoside analogues through a mitochondrial pathway like in a human and the lack of this sufficient fidelity prevented the prediction of the catastrophe (McKenzie et al. 1100). This is not an isolated case study in the history of pharmacology, but rather the representative one. Across all therapeutic areas and development programmes, approximately 90% of drug candidates that clear preclinical animal testing subsequently fail in human clinical trials. Why is that the case? It's either because they prove toxic in ways animal studies did not anticipate or because they fail to produce the efficacy animal data promised (Hay et al. 41). For an industry that spends over two billion dollars to bring a single drug to market, this figure represents a foundational scientific failure, a systematic mismatch between the biology of the model and the biology of the patient (DiMasi et al. 20). It is against this backdrop that new approach methodologies (NAMs): encompassing AI-driven computational models, human organoids, organ-on-chip platforms, and three-dimensional bioprinted tissue must be evaluated. While these NAMs are showcased as promising technologies that do not harm animals (winning the bioethical debate), the most pressing question is whether they are ready to replace animal testing as the primary evidentiary basis for decisions that determine which experimental compounds are administered to humans for the clinical trials. To do this question justice, however, is complicated due to insufficient evidentiary framework because of its recent emergence.

KEYWORDS

NAMs

The scientific case for transitioning away from animal models is stronger than it has ever been. The central failure of animal-based preclinical testing is biological specificity: the human immune system differs from murine and primate immune systems at the level of cytokine expression profiles, T-cell receptor diversity, pathogen recognition pathways, and microbiome composition (Suntharalingam et al. 1019). These determine whether a drug causes catastrophic immune activation in a human while sparing a primate entirely. In the TGN1412 trial of 2006, six healthy volunteers experienced life-threatening cytokine storms within minutes of receiving an anti-CD28 monoclonal antibody that had produced no adverse effects in primate studies. This is the immunological equivalent of fialuridine (Suntharalingam et al. 1018). The science had been wrong in the same systematic way.

Into this breach, new technologies arrive with genuine and documented capabilities. Emulate Inc.'s human liver-on-a-chip detected drug-induced liver injury (DILI), which accounts for 22% of all failed clinical trials, with 87% accuracy, and achieved 100% specificity in flagging safe compounds (Jang et al. 3). The Hartung laboratory at Johns Hopkins has developed an AI prediction tool, trained on over ten million chemical structures and validated against regulatory databases, that predicts toxicological outcomes across nine conventional animal test categories with 91% accuracy, exceeding the inter-laboratory reproducibility of the animal tests themselves (Hartung, "Artificial Intelligence" 812). Brain organoids derived from human induced pluripotent stem cells have successfully recapitulated Zika-induced microcephaly, a condition mouse models could not replicate without direct intracranial injection, and have captured cellular signatures of Alzheimer's disease progression invisible in rodent systems (Lancaster and Knoblich 964). In April 2025, the FDA announced a roadmap to phase out mandatory animal testing requirements for monoclonal antibodies and other biologics, declaring its intention to make animal testing the exception rather than the norm within three to five years ("FDA Announces Plan"). The regulatory momentum, technological promise, and the moral argument should make this ideal for preclinical trials, right?
Wrong.

Disadvantages of NAMs

The field has not yet confronted what I call the “last benchmark problem,” and until it does, the most enthusiastic claims about NAM superiority rest on epistemologically unstable ground. Every validation study in this literature compares a new method's output to the results of animal testing. The organoid's toxicity prediction is judged correct or incorrect relative to what the rat assay produced. The AI model's accuracy is calibrated against decades of accumulated animal test data. This is the benchmark against which the would-be replacements are measured, and it is the same benchmark the field's central premise declares to be unreliable (Rudroff 806).

The circularity is structural. If animal data constitute the gold standard, NAMs can at best match those results, never formally surpass them, and a NAM that diverges from animal findings will be classified as a false positive, even when it is correctly predicting human biology and the animal model was wrong. Conversely, if animal data are systematically unreliable, then 91% agreement with animal test outputs is evidence of alignment with a biased reference. Thomas Hartung, the director of the Johns Hopkins Center for Alternatives to Animal Testing and the field's most prominent scientific advocate, acknowledged precisely this when he stated: “I would not take a decision based on these systems yet” (Hartung, “End of Animals-Only Testing”). This sentence placed in an otherwise optimistic body of advocacy contains the entire problem in eleven words.

The resolution to this dilemma does not lie in building better organoids alone, though that is necessary. It lies in redesigning the validation architecture. The only benchmark that escapes the circle is prospective human outcome data, tracking the Phase I serious adverse event rates and Phase II efficacy signals of drugs approved for investigational use on the basis of NAM packages, compared against matched cohorts approved through conventional animal programmes. This study does not exist. It has not been designed. The FDA's roadmap, for all its ambition, does not require it (“FDA Announces Plan”). Without it, we are asserting the superiority of a new instrument by means of a broken ruler.

Technical Gap

Even setting aside the validation architecture, one technical gap in the current generation of NAMs demands more candor than the advocacy literature has provided. The immune system remains, for all practical purposes, absent from these models. Hartung has observed that there is essentially no human disease in which inflammation plays no role, and no drug toxicity without it (Hartung, "End of Animals-Only Testing"). Systemic immunity is not one organ among others. It is the integrating layer through which every pathological process and every therapeutic intervention is mediated. Current organoids represent, at best, the biology of a developing fetal organ. They lack vascularized networks capable of sustaining circulating immune populations (Clevers 1242). They do not model the dynamic crosstalk between tissue-resident immune cells, circulating lymphocytes, the complement system, and the microbiome. Organ-on-chip systems have begun to address isolated organ-level immune responses, but the systemic picture, the one that matters for oncological immunotherapies, vaccine development, autoimmune disease, and the catastrophic adverse events that hospitalised the TGN1412 volunteers, remains beyond any current platform (Zhou et al.).

This is not a gap that improved protocols will close next year. It is an engineering challenge of fundamental difficulty: vascularisation at scale, with circulating immune populations, integrated across multiple organ compartments, maintained across the weeks or months a regulatory toxicological study requires. The organ-on-chip literature acknowledges this limitation in passing, typically in a final paragraph (Roznik). It should instead be the organizing constraint around which research priorities are set.

RESISTANCE

There is a further obstacle that the scientific literature almost uniformly ignores. Why does this happen? The reason is because it is not a scientific problem at all! Contract research organisations (CROs), the private companies that conduct the majority of preclinical testing on behalf of pharmaceutical sponsors, are structurally opposed to the NAM transition, because animal testing is their highest-margin service (Roznik). These organisations own facilities built for rodent and primate studies, and their client contracts specify animal test timelines and deliverable formats. Their staff are trained in animal handling, not organoid culture.

The FDA's 2025 roadmap encourages, but does not require, NAMs for investigational new drug applications. In the absence of a mandate, CROs face a first-mover disadvantage: every dollar invested in NAM capacity is a dollar withdrawn from infrastructure that clients continue to demand. One former FDA official warned, in analysis immediately following the April 2025 announcement, that a single patient death credibly attributed to a drug that passed NAM screening but would have been flagged by animal testing could set the entire field back by a decade (Roznik). The regulatory door is open. The commercial incentives point in the opposite direction. The field has not reckoned with this asymmetry.

Transitioning

What, then, does a rigorous path forward look like? It requires, first, intellectual honesty about what has been achieved and what has not. The claim that NAMs are more human-relevant than animal models is almost certainly correct in certain domains including single-organ toxicology, species-specific disease modelling, DILI screening and almost certainly premature in others (Rudroff 815). However, the successes of the liver chip or the brain organoid cannot be used to support the argument of replacing animal testing entirely. While the domains of readiness are specific, the domains of unreadiness are equally specific, and they tend to coincide with the domains of immunology, systemic pharmacokinetics, multi-organ toxicity where patient safety is most acutely at risk.

Second, the field requires the prospective validation architecture described above. The FDA should require, as a condition of the NAM pathway, that sponsors submitting IND applications under NAM-primary packages contribute outcome data like Phase I adverse events and Phase II efficacy signals to a centralized registry that will, within a decade, provide the empirical answer to the question the field is currently answering by assertion and instrumenting it for treating the transition itself as the most consequential scientific experiment of the next ten years, and ensuring we capture the data that will reveal whether we got it right (Hartung, "End of Animals-Only Testing").

Third, a new regulatory category for NAM validation is required. I propose this to a significant question that's frequently overlooked: Does this predict human outcome? The Emulate DILI chip's acceptance into the FDA's pilot programme in 2024 was a meaningful step ("New Approach Methodologies"). The next step is a prospective trial in which the chip's predictions, made blind to clinical outcome, are followed until those outcomes are known. That trial, replicated across multiple NAM platforms and therapeutic contexts, constitutes proof of concept. Its absence from every paper reviewed here is telling.

The case for replacing animal testing with AI and organ cultures is, at its moral core, unanswerable.

The suffering of experimental animals is real.
The systematic failures of animal models are real.
The promise of human-derived, human-relevant biological platforms is real.

The FDA has read the science and the political moment and concluded that the direction of travel is irreversible ("FDA Announces Plan"). On all of this, the advocates are right.

But the history of medicine is filled with interventions that were right for the wrong reasons, and right enough to pass available tests while failing in ways no one had thought to examine. The thalidomide disaster was not, in its first iteration, a failure of animal testing. It was a failure of the assumption that animal tests asked the right questions about teratogenicity. We risk repeating that structural error at larger scale if we allow the perfectly reasonable desire to end animal suffering to outpace our capacity to verify that what replaces it genuinely protects the humans whose safety depends on it (Hay et al. 43).

The last benchmark, the prospective human outcome study that would actually close the validation loop, is the scientific debt this field owes. Until it is paid, the answer to whether AI and organ cultures can replace animal testing remains almost certainly yes, but in specific domains, eventually. But we will not know which domains, or when, until we construct the evidentiary infrastructure to find out. That infrastructure is a scientific and institutional choice. Making it is the most important contribution this generation of researchers could offer to the patients who will follow.

REFERENCES

1. Clevers, Hans. "Modeling Development and Disease with Organoids." *Cell*, vol. 165, no. 6, 2016, pp. 1239–1245. doi:10.1016/j.cell.2016.05.082.
2. DiMasi, Joseph A., et al. "Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs." *Journal of Health Economics*, vol. 47, 2016, pp. 20–33. doi:10.1016/j.jhealeco.2016.01.012.
3. "FDA Announces Plan to Phase Out Animal Testing Requirement for Monoclonal Antibodies and Other Drugs." *U.S. Food and Drug Administration*, 10 Apr. 2025, www.fda.gov/news-events/press-announcements/fda-announces-plan-phase-out-animal-testing-requirement-monoclonal-antibodies-and-other-drugs.
4. Hartung, Thomas. "Artificial Intelligence as a Replacement for Animal Experiments in Neurology: Potential, Progress, and Challenges." *Neurology International*, vol. 16, no. 4, 2024, pp. 805–820. doi:10.3390/neurolint16040060.
5. Hartung, Thomas. "The End of Animals-Only Testing Brings Both Opportunities and Challenges." *Johns Hopkins Hub*, 16 July 2025, hub.jhu.edu/2025/07/16/3-questions-organoids-animal-testing-alternatives-thomas-hartung/.
6. Hay, Michael, et al. "Clinical Development Success Rates for Investigational Drugs." *Nature Biotechnology*, vol. 32, no. 1, 2014, pp. 40–51. doi:10.1038/nbt.2786.
7. Jang, Kyung-Jin, et al. "Reproducing Human and Cross-Species Drug Toxicities Using a Liver-Chip." *Science Translational Medicine*, vol. 11, no. 517, 2019. doi:10.1126/scitranslmed.aax5516.
8. Lancaster, Madeline A., and Juergen A. Knoblich. "Organogenesis in a Dish: Modeling Development and Disease Using Intestinal Organoids." *Science*, vol. 345, no. 6194, 2014. doi:10.1126/science.1247125.
9. McKenzie, Rosemary, et al. "Hepatic Failure and Lactic Acidosis due to Fialuridine (FIAU), an Investigational Nucleoside Analogue for Chronic Hepatitis B." *New England Journal of Medicine*, vol. 333, no. 17, 1995, pp. 1099–1105. doi:10.1056/NEJM199510263331702.
10. "New Approach Methodologies (NAMs) Drug Development Pilot Program." *U.S. Food and Drug Administration*, Sept. 2024, www.fda.gov/drugs/drug-approvals-and-databases/new-approach-methodologies-nams-drug-development-pilot-program.
11. Roznik, Katerina. "Organoids, AI, and the FDA's Move Away from Animal Models." *Biotech Blueprint*, 29 May 2025, www.biotechblueprint.com/p/organoids-ai-and-the-fdas-move-away.
12. Rudroff, Thorsten. "Artificial Intelligence as a Replacement for Animal Experiments in Neurology: Potential, Progress, and Challenges." *Neurology International*, vol. 16, no. 4, 2024, pp. 805–820. doi:10.3390/neurolint16040060.
13. Suntharalingam, Ganesh, et al. "Cytokine Storm in a Phase 1 Trial of the Anti-CD28 Monoclonal Antibody TGN1412." *New England Journal of Medicine*, vol. 355, no. 10, 2006, pp. 1018–1028. doi:10.1056/NEJMoa063842.
14. Zhou, Lingrui, et al. "Organoids and Organs-on-Chips: Recent Advances, Applications in Drug Development, and Regulatory Challenges." *Med*, vol. 6, no. 4, 2025. doi:10.1016/j.medj.2025.100667.