"Science is the great antidote to the poison of enthusiasm and superstition."
Adam Smith (1723-1790), Scottish economist and philosopher

Food for Thought ...

The Validation of Regulatory Test Methods – Conceptual, Ethical, and Philosophical Foundations

Thomas Hartung

Johns Hopkins University, Bloomberg School of Public Health and Whiting School of Engineering, Center for Alternatives to Animal Testing (CAAT), Baltimore, MD, USA; CAAT-Europe, University of Konstanz, Germany; Doerenkamp-Zbinden Chair for Evidence-based Toxicology, Baltimore, MD, USA

Received September 27, 2024;

Correspondence:
Thomas Hartung, MD, PhD,
Center for Alternatives to Animal Testing (CAAT),
Johns Hopkins University,
615 N Wolfe St., Baltimore, MD, 21205, USA
(THartun1@jhu.edu)

OPEN ACCESS

ALTEX 41(4), 525-544.
doi:10.14573/altex.2409271

Abstract

Validation establishes the reproducibility and relevance of regulatory test methods, particularly for new approach methods (NAMs) as alternatives to animal testing. While validation concepts provide a framework to assess method suitability, they rarely undergo method-critical assessment. This paper explores the philosophical and ethical foundations of the validation process, drawing from various philosophical traditions and contemporary ethical frameworks. How validation intersects with utilitarian principles, ethics of responsibility, and post-modern critiques is examined, offering a multifaceted perspective on its role in scientific progress and societal values. The paper argues for a paradigm shift in validation, moving beyond traditional animal-based comparisons towards more flexible, fit-for-purpose approaches that embrace emerging technologies and ethical considerations. Key ethical principles guiding NAM validation are discussed, including beneficence, non-maleficence, justice, and respect for animal welfare. Integrating these principles with scientific rigor can create a more holistic validation framework that balances human safety, animal welfare, and technological innovation. By critically examining the philosophical underpinnings of validation, this paper aims to stimulate dialogue on reforming the process to better align with contemporary scientific knowledge, ethical standards, and societal expectations. It calls for a more adaptive, transparent, and ethically grounded approach to validation that can accelerate the adoption of innovative and human-relevant toxicological methods while maintaining scientific integrity and public trust.

Plain language summary

How do we know if new methods for testing chemical safety are reliable and relevant? This process, called validation, is crucial for protecting public health and reducing animal testing. This paper explores the ethical and philosophical ideas behind validation, asking important questions about fairness, animal welfare, and scientific progress. It is argued that current validation methods need updating to keep pace with new technologies and changing social values. By examining different philosophical viewpoints, ways to make validation more flexible, transparent, and ethically sound are suggested. This matters because better validation can lead to safer products, less animal suffering, and more effective environmental protection. The goal is to spark a conversation about how we can improve the way we evaluate new safety testing methods, balancing scientific rigor with ethical considerations and public trust.

1 Introduction

Regulatory authorities need to consider various factors such as hazard assessment, exposure pathways, and public health context when evaluating health and environmental risks and establishing policies. This implies many value decisions: How much risk of possible adverse effects of a substance shall we accept to allow commerce to proceed? What is the value of an animal life? What is the value of advanced methods versus historical use experience? How do we factor uncertainty into these decisions?

This calls, beside a societal/political dialogue, also for an ethical/philosophical discourse. Bhuller et al. (2024, submitted) discuss the complexity of regulatory risk decision-making and the growing importance of incorporating new scientific advancements

into this process; particularly, they address ethical aspects of the regulatory process. As new approach methods (NAMs)[1] and next generation risk assessment (NGRA) frameworks are major tools of change in regulatory frameworks, their responsible introduction and necessary validation need to be discussed. These new methods face various challenges, including the establishment of standardized protocols, the need for rigorous validation studies, and their acceptance by regulatory authorities and stakeholders (Patterson et al., 2021). Bhuller et al. underscore the importance of considering principles for risk decision-making within (the Canadian) regulatory context. This perspective is relevant to the discussion of the ethical and philosophical principles that should guide the validation and implementation of new toxicological test methods, as the article provides a pragmatic analysis of the relevance, importance, and feasibility of such principles in the broader context of health and environmental risk decision-making.

Validation processes are human constructs shaped by cultural, political and subjective factors, not purely objective undertakings. The American anthropologist Clifford Geertz (1926-2006) said "*Man is an animal suspended in webs of significance he himself has spun*." The validation process for new toxicological methods embodies a fundamental tension: It is simultaneously a catalyst for change and a potential obstacle to rapid adoption of new approaches. This dichotomy arises from the dual role of validation as both a necessary quality control measure and a time and resource-intensive hurdle (Box 1).

Box 1: Validation as an accelerator and an obstacle to change of regulatory practices

Validation as an accelerator of change:

a) Rigorous validation lends credibility and legitimacy to new methods. By demonstrating that a new approach is reliable, relevant, and fit-for-purpose, validation helps overcome skepticism and resistance to change. It provides the necessary scientific and regulatory confidence to adopt new methods (Leist et al., 2012; Hoffmann et al., 2017).
b) Formal validation is often a prerequisite for regulatory bodies to accept new methods as replacements for traditional animal tests. Without validation, new approaches may remain confined to research settings. Validation opens the door to regulatory use, which is a major driver of wider adoption and impact (Hoffmann et al., 2017).
c) The validation process involves optimizing and standardizing new methods. This helps ensure reproducibility and consistency across laboratories and applications. By establishing best practices, validation facilitates the transfer and uptake of new methods by the wider scientific community (Stokes and Schechtman, 2007; Judson et al., 2013).
d) The demand for validated alternative methods has spurred intense research and development activity. The validation imperative has focused efforts and resources on creating and refining new techniques to replace animal testing. In this way, validation serves as a key innovation driver in toxicology (Hoffmann et al., 2017).

Validation as an obstacle to change:

a) Validation studies are often lengthy and expensive, requiring multi-lab testing, large datasets, and extensive documentation. This can significantly delay the availability and implementation of new methods. The time and resources required for validation can be a major bottleneck (Hoffmann et al., 2017).
b) Traditional validation frameworks were designed around animal tests as the "gold standard". Strict adherence to these frameworks can disadvantage new approaches that are based on fundamentally different scientific principles. Validation criteria may not adequately capture the unique strengths and limitations of novel techniques, making it harder for them to "prove" themselves against entrenched methods (Hoffmann et al., 2017).
c) Validation processes are often rigid and slow to adapt to scientific advances. Once a method is validated for a specific purpose, it can be difficult to modify or expand its use without re-validation. This can limit the flexibility and iterative improvement of new methods in response to evolving knowledge and needs (Hoffmann et al., 2017).
d) The substantial investment required for validation can discourage ongoing refinement and optimization of new methods. Once a method is validated, there may be reluctance to make changes that would require re-validation, even if they could improve performance or applicability. This can slow the pace of incremental progress (Hoffmann et al., 2017).

Potential ways forward include (1) tiered and flexible validation frameworks (i.e., matching the level of validation to the intended use and risk of the method); (2) proactive validation (i.e., validating promising new approaches in parallel with their development, rather than as a separate, post-hoc process); (3) leveraging computational modeling (i.e., using *in silico* tools and other efficient validation strategies alongside wet-lab testing (Hartung et al., 2024a); (4) collaborative validation efforts (i.e., pooling resources and data across stakeholders to accelerate the process); and (5) adaptive validation frameworks (allowing for iterative refinement and expansion of validated methods as new data emerges). Ultimately, the goal should be to optimize validation as a tool for enabling and accelerating the responsible adoption of new methods, while minimizing unnecessary barriers and delays. This requires ongoing dialogue and innovation in validation practice to keep pace with the rapidly advancing science of toxicology as discussed in this series (Hartung, 2007).

Validation is a tool for systematically establishing credibility. Grinnell's (2009) book *Everyday Practice of Science*[2] provides

[1] https://nc3rs.org.uk/3rs-resources/toxicology-and-regulatory-sciences-bibliography-and-resources

[2] Available at: https://bit.ly/3Y6RfIH

important perspectives to the discussion of scientific credibility and practice. It emphasizes the non-linear, ambiguous nature of actual scientific practice in contrast to the linear "scientific method" often taught. Grinnell argues real science involves intuition and passion as much as objectivity and logic. He discusses how scientists make value judgments throughout the research process about what to study, how to do it, who should fund it, and how to interpret results. Relevant to our discussions here is the process of establishing credibility at the interface of science and society, including research integrity, conflicts of interest, informed consent in human subject research, and the relationship between science and religion. Vigni et al.[3] discuss the fundamentals of scientific credibility in light of the "*erosion of scientists' credibility*". Credibility is established at the interfaces between different social worlds, rather than just within academic science. The paper looks at how scientists establish credibility when engaging in activities that span academia, industry, policy, and public spheres. It focuses on four key dimensions where credibility vectors are being reconfigured: (a) open data practices, (b) science-industry relations, (c) interdisciplinary research, and (d) public engagement and advocacy by scientists. The authors argue that academic publications and disciplinary affiliations remain central and ask for more research on how quantitative metrics (especially the commercialization of academic publishing and the role of private actors) are reshaping scientific credibility. Validation is the answer to this credibility problem. In fact, the information flooding with scientific papers combined with a decline of the quality of the peer-review process makes validation more important than ever.

This article is the first of three parallel articles addressing validation of NAMs. The second article on what can be learned from translational medicine and biomarkers emphasizes how validation can become focused on human relevance (Hartung et al., 2024b): By incorporating clinically validated biomarkers and translational approaches, the validation process can better predict human safety outcomes rather than simply reproducing animal test results. This approach emphasizes the importance of mechanistic understanding and biological pathways that are directly relevant to human physiology. The use of human-derived cells, tissues, and biomarkers in NAMs allows for a more accurate representation of human biology, potentially improving the predictive power of these methods for human toxicity. Additionally, the article highlights how biomarkers can serve as bridges between preclinical and clinical studies, facilitating the translation of findings to human health outcomes. By adopting this biomarker-centric, translational approach to validation, the toxicology field can move towards more human-relevant safety assessment methods, ultimately leading to more effective protection of human health while reducing reliance on traditional animal testing.

The third article emphasizes how artificial intelligence (AI) can make validation more effective (Hartung et al., 2024a). The manuscript proposes "e-validation" as an AI-powered approach to revolutionize and accelerate the validation of new approach methods (NAMs) for toxicology testing. E-validation aims to address key challenges in the current validation process by leveraging AI and machine learning techniques. The approach comprises several integrated modules: smart selection of reference chemicals using clustering algorithms, simulation of validation studies to optimize designs, mechanistic validation to assess biological relevance, and AI-enhanced training to support implementation. A central dashboard coordinates these components. By harnessing AI capabilities, e-validation could dramatically compress traditional decade-long validation timelines while using fewer resources. This AI-driven framework has the potential to finally unlock the promise of human-relevant NAMs and integrated testing strategies by making their validation more efficient, comprehensive, and scientifically robust. Ultimately, e-validation represents a transformative opportunity to overcome the longstanding "validation bottleneck" and accelerate the adoption of innovative toxicological methods.

2 Challenges to the validation process

The principles of validation were shaped by CAAT under the leadership of Alan Goldberg and John Frazier. In 1986, the fourth CAAT Symposium was titled "In Vitro Toxicology: Approaches to Validation" (Goldberg, 1986). Three CAAT/ECVAM/ERGATT (European Research Group for Alternatives in Toxicity Testing) workshops starting in 1990, aka the Amden workshops, developed this further (Balls et al., 1990,1995; Spielmann et al., 1998). Goldberg et al. (1993) summarized the state of the art at the time and later shaped the ECVAM validation process (Worth and Balls, 2004). The author first summarized (in German) the principles of the validation process in an article three decades ago (Hartung and Spielmann, 1995), which might illustrate the struggle of the field with this issue. The validation principles and procedures for alternative methods have not changed since and thus have both challenges and opportunities for improvement:

a) Validation of alternative methods to animal testing aims to determine reproducibility and relevance of the new method compared to animal experiments.
b) There are different validation needs for toxicological tests (consumer safety) versus pharmacological methods (identifying active compounds).
c) The validation process involves several key stages:
   i. Test development
   ii. Evaluation
   iii. Pre-validation
   iv. Validation
   v. Analysis and reporting
d) Pre-validation is an important step to optimize the protocol and check transferability before full validation.
e) Full validation typically involves:
   i. Establishing a management team
   ii. Designing the validation study

[3] https://journals.openedition.org/rac/30500

iii. Conducting a preliminary validation
iv. Conducting the definitive validation (multi-lab, blinded study)

f) Statistical analysis should evaluate validity, reproducibility, precision, sensitivity, specificity, and predictive value of the new method.
g) Final assessment should be done by independent reviewers.
h) Challenges include limited animal data for comparison, species differences, and capturing all aspects of *in vivo* biology *in vitro*.
i) Validation is resource-intensive but necessary for regulatory acceptance. New approaches may be needed for complex *in vitro* methods.
j) Validation should be a shared responsibility between method developers, industry, regulators, and other stakeholders.

Similar processes were developed in the US in 1995 by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)[4,5]. These concepts are the core of OECD Guidance Document 34 (OECD, 2005). In 2004, the modular approach by ECVAM (Hartung et al., 2004) suggested some new concepts and was referenced in OECD GD 34 only one year later.

a) *Modular approach:* The 2004 paper proposes breaking down the validation process into seven independent modules, allowing for more flexibility in the validation process.
b) *Emphasis on information rather than process:* The focus is now on gathering the necessary information to assess test validity, rather than adhering strictly to a predefined process.
c) *Combination of retrospective and prospective validation:* The new approach allows for using existing data (retrospective validation) in combination with new studies (prospective validation) to complete the modules.
d) *Separation of between-laboratory variability and predictive capacity*: These can now be assessed independently, potentially reducing costs and time.
e) *Applicability domain:* Translation of applicability domain, well-established for *in silico* methods, to *in vitro* methods.
f) *Performance standards:* The concept of establishing performance standards for validated tests is introduced, allowing for faster validation of similar "me-too" tests.
g) *Catch-up validation:* This allows for expedited validation of similar tests or improved versions of already validated tests.
h) *Applicability to new technologies:* The modular approach is designed to be more adaptable to new technologies like genomics, proteomics, and *in silico* models.
i) *Periodic reassessment:* The paper suggests creating an index for validity statements, allowing for updates as technologies evolve.
j) *Flexibility in number of laboratories:* Instead of always requiring 3-4 laboratories, the number can be determined on a case-by-case basis.
k) *Emphasis on mechanistic relevance:* There is increased focus on understanding and describing the mechanistic basis of tests.

These changes represent a more flexible, efficient, and adaptable approach to validation compared to the traditional approach, which is a more rigid, step-wise process. Hoffmann and Hartung (2006a) demonstrated how validation studies could be designed more efficiently, modeling on the validation of the EPISKIN test for skin corrosion.

Test method performance standards are powerful tools for streamlining the validation process of NAMs by providing predefined benchmarks that novel assays must meet or exceed. These standards include criteria such as sensitivity, specificity, reproducibility, and applicability, which serve as reference points for evaluating the performance of new tests. By establishing clear and objective performance standards, the need for extensive comparative testing against traditional animal models can be reduced, accelerating the validation timeline. Performance standards facilitate a more efficient assessment by focusing on whether a new method performs within the acceptable range of a validated "gold-standard" assay, thus avoiding repetitive validation studies for methods that demonstrate similar or superior performance. This approach not only enhances the reproducibility and reliability of the validation process but also provides a pathway for validating groups or variations of related methods – often referred to as "me-too" tests – under a common set of criteria. Consequently, performance standards support the faster adoption of NAMs in regulatory frameworks, reduce resource expenditure, and encourage innovation by allowing researchers to focus on the unique capabilities of their methods rather than exhaustive validation procedures. They help us to not validate ourselves to death!

Already in 2007, a paper in this Food-for-Thought … series (Hartung, 2007) addressed validation: This paper described several challenges with the validation process for alternative methods in toxicology, including:

a) Lack of human data as a reference point, forcing comparison to animal tests which may not be accurate.
b) The precautionary principle, leading to many false positives in animal tests.
c) Lack of standardized protocols for animal tests used as reference points.
d) Difficulty standardizing alternative methods while allowing flexibility.

Challenges in validating integrated testing strategies rather than single replacement tests:

a) Lack of post-validation monitoring of methods.
b) Insufficient focus on mechanistic relevance of alternative methods.

The paper proposes several novel concepts to address these challenges:

a) Using meta-analysis and weight-of-evidence (WoE) approaches to combine data from multiple sources. Hartung (2010) expanded this to suggest evidence-based toxicology as a toolbox for validation (see below).

4 https://ntp.niehs.nih.gov/sites/default/files/iccvam/docs/about_docs/validate.pdf
5 https://ntp.niehs.nih.gov/sites/default/files/2024-03/VWG_Report_27Feb2024_FD_508.pdf

b) Defining performance standards to assess equivalence of similar methods.
c) Focusing more on human relevance and mechanistic basis rather than just concordance with animal data.
d) Allowing for both prospective and retrospective validation approaches.
e) Developing tools to compose and validate integrated testing strategies.
f) Implementing post-validation surveillance and user feedback mechanisms.
g) Formalizing the assessment of the scientific/mechanistic basis of methods.
h) Taking a more flexible, modular approach to validation rather than following a rigid process.
i) Moving towards an evidence-based toxicology paradigm with validation as a key component.

Our 2010 article "*Evidence-Based Toxicology – the Toolbox of Validation for the 21st Century?*", written in the context of a symposium organized by CAAT on *21st Century Validation for 21st Century Toxicology*, developed some key novel concepts proposed for the validation process in toxicology which include:

a) Moving away from traditional validation against animal tests, towards more flexible approaches that can accommodate new complex methods and paradigm shifts.
b) Incorporating concepts from evidence-based medicine (EBM), particularly EBM approaches for evaluating diagnostic tests. This includes:
   i. Using systematic reviews and meta-analysis
   ii. Weighing evidence by quality scores
   iii. Using statistical tools like likelihood ratios and ROC curves
   iv. Considering pre- and post-test probabilities
c) Focusing more on scientific relevance and mechanistic understanding rather than just reproducibility and predictivity of animal test results.
d) Validating pathways of toxicity (PoT) rather than just individual tests or assays (Kleensang et al., 2014).
e) Developing approaches to validate integrated testing strategies (ITS) rather than just single methods.
f) Incorporating prevalence and moving to predictive values rather than just sensitivity/specificity.
g) Defining adversity based on PoT signatures rather than just apical endpoints.
h) Moving towards probabilistic risk assessment with confidence intervals rather than binary classifications (Maertens et al., 2022, 2024).
i) Using retrospective validation approaches that evaluate existing data rather than just prospective studies.
j) Developing quality assurance approaches for new technologies like omics and *in silico* methods.
k) Considering factors like selection bias and spectrum bias that affect test performance in practice.

The overall goal is to develop more flexible, scientifically based validation approaches that can evaluate complex new methodologies while still ensuring their reliability and relevance for human toxicology. This represents a shift from traditional validation paradigms towards evidence-based toxicology (Hartung, 2010).

John Frazier first suggested using mechanistic toxicology for validation (Frazier, 1994) but there was no follow-up. However, the need to modernize the validation guidance was perceived, which led to a respective ongoing OECD project. The article "*Mechanistic Validation*" by Hartung et al. (2013a) argues for a redefinition of the validation process to include mechanistic validation as a core component, shifting from traditional phenomenological methods to those grounded in mechanistic and systems-based approaches. Traditional validation, which focuses on reproducibility and predictive capacity, often neglects the scientific basis or mechanistic understanding of test systems (Fig. 1). In the new paradigm, mechanistic validation involves demonstrating the causality and relevance of biological mechanisms that a test intends to measure, using tools like the Bradford Hill criteria and bioinformatics for complex systems analysis. This approach aligns with initiatives such as Tox21 and the Human Toxome Project (Bouhifd et al., 2015), which advocate for pathway-based methodologies. The article suggests that validation can move beyond simple comparisons with traditional animal models, focusing instead on understanding how and why a test works, and thereby increasing its predictive value and relevance. This redefined framework proposes a pathway to overcome limitations in current validation processes and aligns with the broader goals of evidence-based toxicology and the evolving landscape of regulatory science.

The validation of new approach methods (NAMs) as alternatives to animal testing faces several significant challenges that hinder their development, acceptance, and implementation. These challenges arise from a combination of scientific, regulatory, and practical factors that need to be addressed to facilitate the successful adoption of NAMs (Hartung, 2007; Leist et al., 2012).

a) *Insufficient engagement with end users during NAM development:* One major issue is that NAMs have often been developed without adequate consultation and input from key end users, such as regulatory agencies and industry stakeholders (Leist et al., 2012). This lack of communication has led to the creation of NAMs that do not fully meet the specific testing requirements and needs of these end users. Consequently, even if a NAM is technically validated, it may not be accepted or widely used by regulators and industry because it does not provide the necessary information or fit into their existing testing frameworks. To overcome this challenge, NAM developers need to engage with end users early in the development process to ensure that the NAMs are designed to meet their specific needs and can be seamlessly integrated into regulatory and industry practices (ICCVAM, 2018).
b) *Limitations of the traditional validation process:* The current validation process for NAMs, which is largely based on OECD GD 34 (OECD, 2005), has several limitations that hinder the efficient and effective validation of new methods. Firstly, the process is often lengthy and resource-intensive, requiring significant time and financial investment. This can delay the availability of promising NAMs and discourage their development. Secondly, although OECD GD 34 allows for a modular approach to validation, which could provide flexibility and efficiency, this approach has not been fully utilized in practice (Hartung et al., 2004). Instead, a more rigid and comprehensive

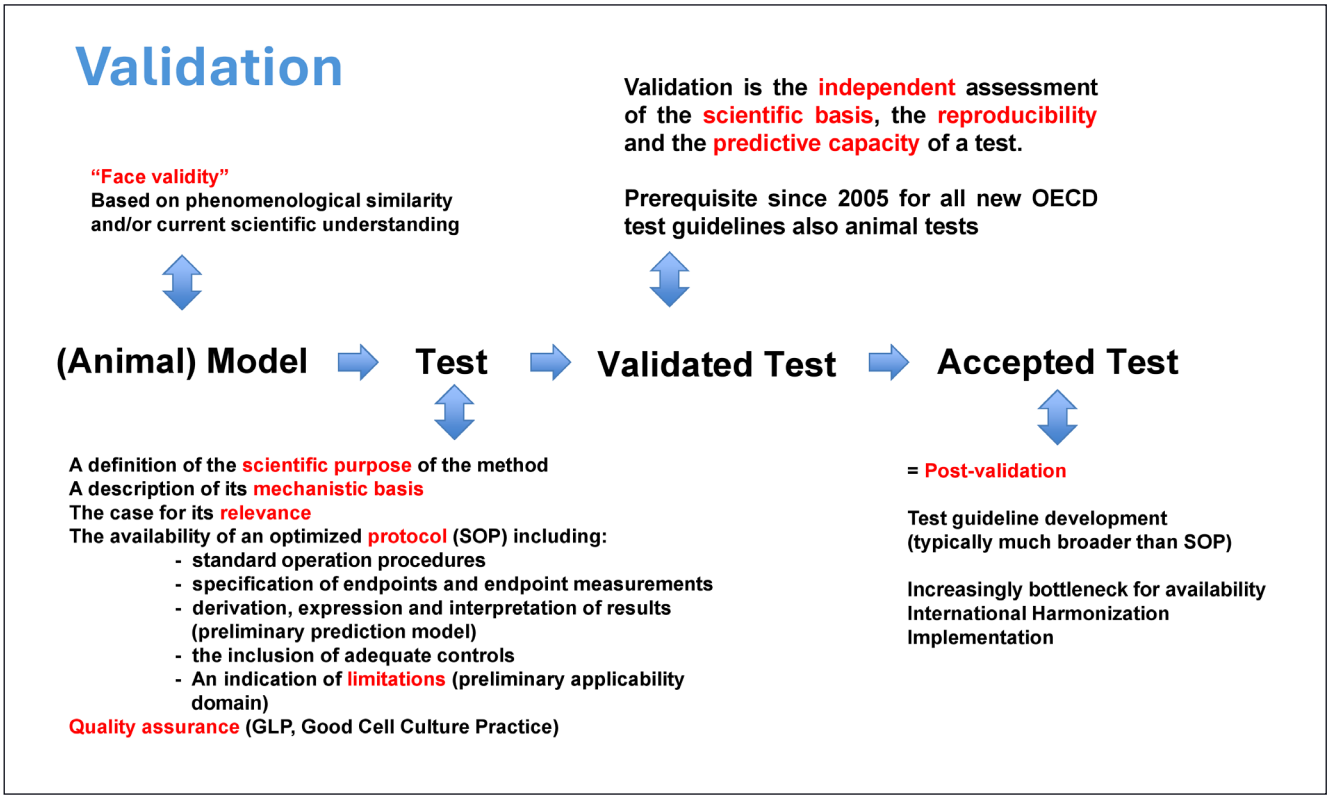

Fig. 1: The current validation framework

The current framework almost exclusively aims to replace a given animal test one-to-one. The animal test's validity is taken at face value. Key to the process is a strong test definition, which freezes the test in time. The validation process has a strong focus on reproducibility and prediction of the results of the animal test (little of the scientific basis, see Hartung et al., 2013a). The post-validation process (Bottini et al., 2008) is still not well-defined.

validation process has often been applied, which may not be necessary or appropriate for all types of NAMs. Lastly, OECD GD 34 was developed before the advent of many modern technologies and approaches, such as high-throughput screening and computational modeling. As a result, it does not fully address the specific considerations and requirements for evaluating these novel methods. To address these limitations, a more flexible, efficient, and technology-agnostic validation framework is needed (Parish et al., 2020).

c) *Challenges in comparing NAMs to animal tests:* Validating NAMs by comparing them to traditional animal tests can be problematic for several reasons. Firstly, many animal tests have not been thoroughly validated themselves for human relevance and reproducibility (Leist and Hartung, 2013). This means that the animal data may not be a reliable or appropriate standard against which to evaluate the performance of NAMs. Secondly, animal tests often exhibit low reproducibility, both within and between laboratories (Browne et al., 2019). This variability makes it difficult to establish clear performance standards for NAMs and may lead to the rejection of promising methods that do not meet an arbitrarily high threshold of concordance with animal data. In some cases, NAMs are being held to a higher standard of reproducibility and reliability than the animal tests they are intended to replace (Hoffmann et al., 2008). Thirdly, for complex endpoints such as developmental neurotoxicity or systemic toxicity, a battery of NAMs may be needed to replace a single animal test (Hartung et al., 2013b; Rovida et al., 2015; Caloni et al., 2022). In these cases, it is challenging to determine how to validate the individual NAMs and the overall testing strategy. Should each NAM be validated independently against the animal test, or should the performance of the entire battery be evaluated? These questions highlight the need for new validation approaches that focus on the scientific validity and human relevance of NAMs, rather than their ability to reproduce animal data (Clippinger et al., 2021).

d) *Insufficient post-validation evaluation and refinement:* Another challenge in the validation of NAMs is the lack of systematic post-validation evaluation and refinement. Once a NAM has been validated and accepted for use, there is often limited follow-up to gather feedback on its performance in real-world applications and to identify areas for improvement (Leist et al., 2012). This is problematic because the validation process provides only a snapshot of a method's performance at a specific point in time, under specific conditions. As new scientific knowledge emerges and experience with the method accumulates, it may become clear that the NAM needs to be refined

or updated to maintain its relevance and reliability. Without a mechanism for ongoing evaluation and refinement, NAMs may become outdated or fail to meet evolving regulatory and industry needs. To address this challenge, a more iterative and adaptive validation process is needed, one that includes provisions for post-validation monitoring, feedback, and continuous improvement (Bottini et al., 2008).

e) *Over-reliance on concordance with animal data:* A fundamental challenge in the validation of NAMs is the heavy reliance on comparing their performance to animal data, rather than focusing on their scientific mechanistic basis and human relevance. This raises the fundamental question, what to compare to, i.e., the point of reference (Hoffmann et al., 2008). This goes far beyond the reference material question of the validation of analytical methods (Bower et al., 2014) as also the relevance of a method is being assessed. If we choose the traditional animal method as point of reference, we will get an estimate of how well the animal reaction is predicted, not whether humans are protected. This approach assumes that animal tests are the "gold standard" for assessing chemical safety and that NAMs must replicate their results to be considered valid. However, this assumption is increasingly being questioned, as animal tests have many limitations and may not accurately predict human responses (Hartung, 2010). By focusing solely on concordance with animal data, the validation process may overlook the potential of NAMs to provide more human-relevant and mechanistically informative data. Moreover, this approach perpetuates the use of animal tests and hinders the development of truly innovative and transformative NAMs. To overcome this challenge, a paradigm shift is needed in the validation process, one that prioritizes the assessment of a NAM's scientific validity, mechanistic basis, and human relevance, rather than its ability to match animal data (Clippinger et al., 2021).

f) *Need for a more flexible and efficient validation framework:* To address the aforementioned challenges and facilitate the successful development and implementation of NAMs, a more flexible and efficient validation framework is needed. This framework should be adaptable to different types of NAMs, including high-throughput screening, computational modeling, and multi-parametric assays. It should also be able to accommodate the validation of individual NAMs and integrated testing strategies, depending on the specific context and requirements (Parish et al., 2020). The framework should emphasize the assessment of a NAM's fitness-for-purpose[6,7], mechanistic relevance, and human predictivity, rather than focusing solely on concordance with animal data (Hartung, 2007). Noteworthy, most validation activity was done in the context of replacement alternatives while reduction and refinement were less in the focus of formal validation. The upcoming validation of virtual control groups (Steger-Hartmann et al., 2020; Golden et al., 2024) in the IMI-2 VICT3R project[8] will be an interesting challenge in this respect.

Furthermore, the framework should include provisions for staged validation, where NAMs can be evaluated and accepted for specific uses or contexts, rather than requiring a one-size-fits-all approach. This would allow for the gradual building of confidence in a NAM and its progressive implementation in regulatory and industry practices. By adopting a more flexible and efficient validation framework, the development and acceptance of NAMs can be accelerated, leading to a faster transition away from animal testing and towards more human-relevant and mechanistically based approaches to chemical safety assessment (ICCVAM, 2018).

In 2022, van der Zalm et al. saw the following challenges for validation:

- Traditional validation processes are lengthy, expensive, and cumbersome.
- There has been an overreliance on comparing NAM results to animal test data, which may have limited human relevance.
- Many regulatory requirements were written for animal tests and are not easily adapted for NAMs.
- There is often a long delay between NAM development and regulatory acceptance.
- Inter-laboratory reproducibility studies (ring trials) are resource-intensive.

With the overall goal to establish a more efficient, flexible process focused on human relevance to accelerate regulatory acceptance of NAMs and to address these challenges, the paper proposes several concepts for NAM validation:

- Focus on human biological relevance rather than concordance with animal data.
- Evaluate fitness-for-purpose based on regulatory needs rather than one-to-one replacement of animal tests.
- Use reproducibility of animal tests to set realistic performance benchmarks for NAMs.
- Emphasize transparent description of NAM strengths and limitations.
- Allow for flexible, fit-for-purpose validation approaches rather than rigid processes.
- Incorporate early engagement between regulators and NAM developers.
- Use a framework with 5 key elements: fitness-for-purpose, human biological relevance, technical characterization, data integrity/transparency, and independent review.
- Focus on mechanistic understanding and ability to support health-protective decisions.
- Allow for qualitative comparisons to animal data when quantitative comparisons are not feasible.

Mondou et al. (2021) adds several key insights to the discussion on validating and adopting NAMs for chemical risk assessment:

6 ICCVAM defined fit-for-purpose as "*a process based on scientifically sound principles by which the relevance and reliability of a particular method or process are established for a specific purpose (OECD GD34)*"; "*Focus on purpose, as opposed to demonstrating equivalence to animal based results*"; "*Acknowledge that no assay is perfect, recognize that most will be used in weight of evidence approach.*"

7 https://www.toxicology.org/groups/rc/ncac/docs/091815-7CASEY.pdf (accessed 02.09.2024)

8 https://www.vict3r.eu

a) It identifies two key priorities for facilitating NAM validation internationally:
   i. Developing common data collection, reporting, and sharing procedures
   ii. Improving knowledge about new test methods among regulators
b) It suggests the need for a common regulatory science infrastructure, including:
   i. International regulatory dialogues
   ii. Large-scale research collaborations
   iii. Coordinated innovation in technological tools, validation discourse, and regulatory procedures
c) It highlights the importance of building trust across multiple stakeholder groups (labs, regulators, industry, public) by:
   i. Agreeing on validation requirements for specific uses
   ii. Standardizing how results are communicated and measured
d) It proposes the need for a global "orchestrator" to inspire voluntary cooperation among diverse organizations to address shared validation goals.
e) It identifies challenges in NAM validation, including:
   i. Lack of consensus on validation approaches
   ii. Complexity of interpretation
   iii. Need for standardization
f) It suggests potential frameworks to facilitate NAM adoption, like integrated approaches to testing and assessment (IATA) and adverse outcome pathways (AOPs).
g) It emphasizes the need for a flexible, fit-for-purpose approach to validation that considers the context of use.
h) It highlights the tension between the need for global standards and the challenges of international harmonization in a complex governance landscape.

Overall, the paper provides a forward-looking perspective on the policy challenges and potential solutions for accelerating the international validation and adoption of NAMs in chemical risk assessment.

The US National Academy of Sciences report titled "*Building Confidence in New Evidence Streams for Human Health Risk Assessment: Lessons Learned from Laboratory Mammalian Toxicity Tests*" (NASEM, 2023)[9] introduces a structured framework for evaluating the scientific confidence of NAMs, offering both a theoretical basis and practical recommendations for evaluating NAMs: The report argues that the traditional concept of validation is no longer sufficient for NAMs due to their complexity and diverse applications. Instead, it introduces the term "scientific confidence framework," which broadens the scope of validation to include aspects like reliability, relevance, fitness-for-purpose, and transparency. This aligns with our argument here for more flexible and adaptive validation paradigms for NAMs. This report highlights the need for a shift from traditional validation processes to broader frameworks that consider scientific confidence in NAMs:

a) *Fit-for-purpose validation:* The report emphasizes a "fit-for-purpose" validation approach, considering the intended context of use of the NAMs, rather than following a rigid validation protocol. This is particularly relevant to our call for aligning validation processes with the specific goals and applications of NAMs, such as hazard identification, screening, or dose-response assessment.
b) *Framework for addressing internal and external validity:* The report differentiates between internal and external validity for NAMs, defining specific evaluation domains like biological relevance, predictive capacity, and concordance. These criteria can serve as robust benchmarks for determining how well NAMs can replace or complement traditional animal models. This could help in establishing a comprehensive framework for evaluating the scientific credibility of NAMs.
c) *PECO (population, exposure, comparator, outcome) statements:* The report advocates for the use of PECO statements to define the scope and context of use for each NAM. This systematic approach could help clarify the utility of NAMs in specific research or regulatory contexts and assist in harmonizing validation criteria across different methodologies and applications.
d) *Integration of mechanistic and traditional evidence:* It discusses how mechanistic evidence, derived from NAMs, can be integrated with traditional data to support human health risk assessment. This integration is crucial when discussing the transition from animal-based models to NAMs and can provide a pathway for achieving regulatory acceptance.
e) *Structured evaluation of confidence:* The report provides a structured approach for evaluating confidence in NAMs using criteria such as transparency, documentation, and systematic assessment of internal and external validity. This method can help standardize how new methods are validated and accepted.

Interestingly, stakeholders in the OECD process recently published a brainstorming paper (Gourmelon et al., 2024) mentioning the following challenges to the validation process:

– Significant resources are needed, which are often underestimated.
– Validation is unappealing to the scientific community, as it does not usually result in publications.
– Lack of funding makes it challenging for academic laboratories to participate.
– There is a disconnect between academic method developers and deployment of new technologies.
– The process requires scientific and technical rigor, where shortcuts can be counterproductive.
– Lack of knowledge in Validation Management Teams on how to design and conduct studies.
– Reluctance to invest in validation of innovative methods without clear indications of regulatory use.
– Difficulty in identifying suitable reference chemicals for some endpoints (e.g., endocrine disruptors).
– Challenges in transferring methods to laboratories with limited proficiency in specific techniques.

[9] https://nap.nationalacademies.org/read/26906/chapter/7

In conclusion, the validation of NAMs faces several significant challenges that need to be addressed to enable their successful development, acceptance, and implementation. These challenges include insufficient engagement with end users, limitations of the traditional validation process, difficulties in comparing NAMs to animal tests, insufficient post-validation evaluation and refinement, over-reliance on concordance with animal data, and the need for a more flexible and efficient validation framework. By acknowledging and addressing these challenges, the scientific community can work towards creating a more effective and streamlined validation process that facilitates the adoption of NAMs and advances the field of toxicology towards a more human-relevant and mechanistically based approach to chemical safety assessment.

3 Philosophical underpinning of evidence-based approaches

The author has been a strong proponent of evidence-based approaches to toxicology (Hoffmann and Hartung, 2006b; Hartung, 2009, 2023c; Stephens et al., 2013; Krewski et al., 2022; Hartung and Krewski, 2022; Hartung and Tsaioun, 2024) and validation (Hartung, 2010). Noteworthy, there are philosophical aspects and foundations associated with evidence-based approaches. The core philosophy behind evidence-based approaches is empiricism, which emphasizes the importance of observable evidence in forming knowledge and making decisions. While not entirely aligned, evidence-based approaches share some similarities with logical positivism, particularly in their emphasis on verifiable observations and scientific methods. They also relate to Karl Popper's falsificationism – the idea that scientific theories should be testable and potentially falsifiable is central to evidence-based approaches. His philosophy of critical rationalism, which emphasizes critical thinking and the importance of attempting to disprove hypotheses, has also influenced evidence-based approaches. Thomas Kuhn's paradigm shifts concept provides an understanding of how evidence is interpreted and how consensus is formed in evidence-based fields; the resonance with toxicology has been discussed twice (Hartung, 2008; Hartung and Tsatsakis, 2021). The use of Bayesian reasoning in updating beliefs based on new evidence is increasingly important in evidence-based approaches. The philosophical tradition of pragmatism, particularly its emphasis on practical consequences, aligns well with evidence-based approaches. The philosophical view that theories and models should be judged primarily by their predictive power and practical usefulness (instrumentalism) aligns with many evidence-based practices. Broader questions in the philosophy of science, such as the nature of scientific evidence, causality, and inference, are relevant to evidence-based approaches. Similarly, epistemology addresses questions about the nature of knowledge, justification, and belief, which are central to understanding how evidence informs knowledge and decision-making. The social epistemology branch of philosophy, which examines social dimensions of knowledge, is relevant to understanding how evidence is produced, disseminated, and used in social contexts. These philosophical underpinnings provide a framework for understanding the strengths and limitations of evidence-based approaches. They also inform ongoing debates about the nature of evidence, how it should be gathered and interpreted, and how it should be applied in various fields such as medicine, policy-making, and scientific research.

It is worth noting that while evidence-based approaches are generally viewed positively, they are not without critique. Some philosophers and scholars have raised concerns about the potential limitations of these approaches, such as the risk of overlooking important but difficult-to-measure factors or the challenges in applying population-level evidence to individual cases.

The ethical and philosophical underpinnings of evidence-based approaches, including evidence-based toxicology (EBT), are rooted in several key principles:

a) *Empiricism:* Evidence-based approaches are fundamentally grounded in empiricism, the idea that knowledge should be based on observable, measurable evidence rather than intuition, authority or dogma. This aligns with the scientific method and the idea that claims should be testable and falsifiable.
b) *Objectivity:* EBT strives for objectivity in evaluating evidence, minimizing bias and subjective judgment. This is reflected in the use of explicit, systematic methods for identifying, appraising, and synthesizing evidence. Philosophically, this embodies the ideal of science as a dispassionate, unbiased pursuit of truth.
c) *Transparency:* Evidence-based approaches prioritize transparency in methods, data, and reasoning. Systematic reviews, for example, have extensive reporting guidelines to ensure reproducibility. This commitment to transparency reflects values of openness, honesty, and accountability in science.
d) *Epistemic humility:* EBT recognizes the limitations and uncertainties in scientific knowledge. It acknowledges that individual studies may be flawed or context-dependent, and that even rigorous evidence synthesis may not yield definitive answers. This epistemic humility aligns with the fallibilist view that scientific knowledge is always provisional and subject to revision based on new evidence.
e) *Pragmatism:* Evidence-based approaches aim to provide practical guidance for real-world decision-making, even in the face of imperfect evidence. The use of systematic reviews and meta-analyses to synthesize the best available evidence reflects a pragmatic philosophy of science in service of action.
f) *Utilitarianism:* The ultimate goal of EBT is to make toxicological risk assessment and decision-making more evidence-based in order to better protect human health and the environment. This reflects a utilitarian ethical framework, aiming to maximize overall societal benefit and minimize harm.
g) *Duty of care:* Applying the best available scientific evidence in toxicological assessments and regulations can be seen as an ethical duty of care towards the public and the environment. EBT provides a framework to uphold this duty rigorously and transparently.

EBT can help to reform the validation process for regulatory test methods in several key ways:

a) *Expanding the evidence base:* Traditional validation heavily relies on prospective ring trials, which can be resource-intensive and time-consuming. EBT encourages the incorporation of a broader evidence base, including retrospective analyses of existing data, mechanistic studies, and human/clinical data when available. This could make the validation process more efficient and informative.
b) *Systematic review of validation studies:* Applying systematic review methodology to the validation process itself could improve the quality and reliability of validation conclusions. Systematically identifying, appraising, and synthesizing all relevant validation studies for a test method could provide a more comprehensive and objective assessment of its performance and fitness-for-purpose.
c) *Meta-analysis of validation data:* When multiple validation studies are available, EBT advocates for the use of meta-analysis to quantitatively synthesize their results similar to the modular approach (Hartung et al., 2004) discussed earlier. This could provide more precise and reliable estimates of a test method's performance metrics (e.g., sensitivity, specificity) compared to relying on individual studies alone.
d) *Weight-of-evidence evaluations:* EBT frameworks could be used to transparently weigh and integrate different lines of evidence in the validation process. This could include considering factors such as the quality and relevance of individual validation studies, the consistency of results across studies, and the strength of mechanistic evidence supporting the test method (see discussion of our earlier paper in this series (Linkov et al., 2015) below).
e) *Probabilistic performance measures:* EBT encourages the use of probabilistic measures such as predictive values, likelihood ratios, and Bayesian posterior probabilities, which account for the prevalence of the toxicity endpoint in the population of substances tested. Incorporating these measures into the validation process could provide a more realistic and informative assessment of a test method's performance in practical use.
f) *Continuous evidence updating:* EBT recognizes that scientific knowledge is always evolving. Establishing processes for the continuous updating of validation conclusions as new evidence emerges could help ensure that regulatory test methods remain current and fit-for-purpose over time.
g) *Transparency and stakeholder engagement:* EBT principles of transparency and stakeholder engagement could be applied to make the validation process more open and inclusive. This could involve the publication of detailed validation protocols, results, and decision-making processes, as well as opportunities for input from a wide range of stakeholders.

WoE is unfortunately a somewhat muddy term. It is often used to describe situations where different incomplete pieces of evidence are pragmatically lumped together. In contrast, Linkov et al. (2015) discuss very rigorous WoE approaches embracing decision theory, which actually relate to EBT and the validation process in several key ways:

a) WoE is an approach that integrates individual lines of evidence to form a conclusion and has been used in the process of validating new and alternative test methods. However, the article argues WoE is currently at a crossroads – while some efforts aim to formalize WoE methodologies, others criticize it as too vague and subjective.
b) EBT advocates for more systematic, objective, and quantitative approaches to evaluating toxicological evidence, drawing inspiration from evidence-based medicine. This aligns with calls to make WoE more rigorous and less reliant on subjective expert judgment.
c) The article proposes that Bayesian statistics, which were part of the original conception of WoE in the 1960s, could enhance the information base and rigor of WoE applications in the context of EBT. Some existing WoE approaches utilizing Bayesian methods are highlighted as examples.
d) Multi-criteria decision analysis (MCDA) is suggested as a suitable proxy for Bayesian analysis in WoE when data limitations preclude formal statistical modeling. MCDA can facilitate the systematic and transparent synthesis of multiple sources of evidence.
e) Integrating EBT principles and tools like systematic review, meta-analysis, Bayesian statistics, and MCDA into WoE could improve the validation process for new test methods by making it more comprehensive, objective, quantitative, and transparent.

In summary, Linkov et al. (2015) positions rigorous, quantitative WoE approaches, aligned with EBT principles, as a potential path forward for enhancing the validation of new toxicological methods in a way that is systematic and scientifically robust. Reforming WoE using EBT tools could increase regulatory and public acceptance of the method validation process. In general, integrating EBT principles and methodologies into the validation process could make it more efficient, comprehensive, objective, informative, and transparent (Hartung, 2010). This could ultimately lead to the more timely adoption of innovative and scientifically sound test methods for regulatory use, while ensuring their reliability and relevance for decision-making.

4 Borrowing from some philosophers

Already Aristotelian *virtue ethics* could be seen as a guiding principle. They focus on the moral character of individuals and institutions rather than rules or consequences. In the context of validation, a virtue ethics lens would emphasize the importance of scientific integrity, transparency, objectivity, and public-spiritedness as key virtues that should guide the conduct of validation studies and decision-making. Several modern philosophical frameworks and concepts can be applied to illuminate challenges and opportunities in the validation of new toxicological test methods:

a) *Thomas Kuhn's paradigm shifts:* Kuhn's theory of scientific revolutions suggests that science progresses through paradigm shifts, in which the dominant set of theories, methods, and assumptions is replaced by a new paradigm. The ongoing transition from animal-based to alternative methods in toxicology can be seen as a paradigm shift. Validation plays a key role in this transition but may also be met with resistance from the existing paradigm.

b) *Karl Popper's falsificationism:* Popper argues that the key criterion for a scientific theory or method is its falsifiability – the possibility of it being proven wrong by empirical evidence. From this perspective, validation studies should be designed to rigorously challenge and attempt to falsify new test methods rather than confirm their performance. Methods that withstand such scrutiny are more scientifically robust.
c) *Jürgen Habermas' discourse ethics:* Habermas' theory emphasizes the importance of open, inclusive, and rational dialogue in resolving ethical and political disputes. Applied to validation, this suggests the need for transparent, participatory processes that engage all relevant stakeholders and consider multiple perspectives. The goal is to reach consensus on the validity and acceptability of new methods through reasoned argumentation.
d) *The precautionary principle:* It is important to note that the precautionary principle has roots in environmental policy and public health as much as in philosophy. This principle holds that when an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause-and-effect relationships are not fully established scientifically. The German philosopher Hans Jonas is credited with laying the philosophical groundwork for the precautionary principle in his book *The Imperative of Responsibility* (1984). Jonas emphasizes the need for a new ethic of long-term responsibility in the face of modern technology's potential for catastrophic harm. In the context of validation, this would suggest erring on the side of caution and requiring a high burden of proof for new methods to be accepted as replacements for established safety tests.
e) *Utilitarianism:* This ethical framework, associated with philosophers like Jeremy Bentham and John Stuart Mill, or in the context of animal rights with Peter Singer, holds that the morally right action is the one that produces the greatest good for the greatest number. From this perspective, the validation process should prioritize test methods that maximize overall social benefits (e.g., improved public health protection, reduced animal use) while minimizing social costs (e.g., economic impacts on industry).

These philosophical perspectives offer different, sometimes conflicting, insights into the validation process. Some will be discussed in the following. Engaging with a range of philosophical frameworks can help to elucidate the ethical, epistemological, and political dimensions of test method validation and inform more robust, inclusive, and socially responsive approaches. Ultimately, validation should aim to strike a balance between scientific rigor, practical feasibility, ethical obligations, and societal values.

In the following, some perspectives that have influenced the author's thinking are discussed in more detail.

4.1 Kuhn's scientific paradigm change (revolution) as the goal

We discussed earlier (Hartung, 2008; Hartung and Tsatsakis, 2021) that this resonates with toxicology:

a) Kuhn proposed that science progresses through cycles of normal science, crisis, and revolution, leading to a paradigm shift. Our article argues that toxicology is currently moving from a period of crisis, where the old animal testing paradigm is increasingly challenged by anomalies and limitations, towards a revolutionary phase where NAMs are gaining traction.
b) According to Kuhn, anomalies that cannot be explained by the current paradigm accumulate during normal science, eventually leading to a sense of crisis. The article lists several such anomalies for toxicology, including the huge animal testing needs of REACH, new types of products needing assessment like biologics and nanomaterials, the drying pharmaceutical pipeline, legislations pushing for alternatives, etc. These anomalies challenge the adequacy of the traditional animal testing paradigm.
c) Kuhn argued that a new paradigm emerges in response to crisis, which initially has few supporters but eventually gains ground as it is explored and advocated. This matches the trajectory of NAMs in toxicology over recent years – from isolated efforts to an increasingly supported vision for transforming the field.
d) Ultimately for Kuhn, scientific revolutions lead to scientists "*working in a different world*" – a transformed perspective on their field. The article suggests toxicology is at the cusp of such a revolutionary reorientation around 21st century paradigms and methods.

In summary, Kuhn's model provides a useful perspective to understand the current state and dynamics of toxicology – the accumulating challenges to the old paradigm, emergence of a new paradigm, and the complex evidentiary and social processes involved in the field's transformation. This highlights that validation of new methods is taking place within a broader revolution in toxicology.

New approaches are at the center of the paradigm change in toxicology, and their validation is the entry port to change to regulatory practice. Kuhn's framework applied to the area of NAM validation suggests that validation efforts could serve as an important catalyst and mechanism for the paradigm shift in toxicology:

a) *Validation as a response to anomalies:* The intensive focus on validating new toxicological methods over the last few decades can be seen as a response to the accumulating anomalies and limitations of animal-based methods. Validation projects are often motivated by the recognition that traditional approaches are inadequate for evolving scientific and regulatory needs. In this sense, validation is driven by the "crisis" in the old paradigm.
b) *Validation as a way to improve new methods:* Kuhn noted that supporters of a new paradigm work to improve and explore its possibilities. Validation serves exactly this purpose for new toxicological methods – it involves optimizing protocols, demonstrating reproducibility, and characterizing performance to establish the credibility and applicability of the new approaches. Validation helps advance new methods from initial concepts to mature, reliable tools.
c) *Validation to change perceptions:* Kuhn emphasized that paradigm choice is not based just on scientific evidence but involves subjective and social factors in the scientific community. Similarly, the adoption of new toxicological methods is shaped by more than just strict validation data, but also changing stakeholder perceptions, values, and comfort levels.
d) *Validation to redefine the field:* For Kuhn, the new and old paradigms are often incommensurable – they define the field

differently. This is reflected in the tension between traditional animal tests as the "gold standard" and the fundamentally different assumptions and approaches of next generation, human biology-based methods. Validation has to evolve to accommodate this shift.

e) *Validation as a means of persuasion:* For Kuhn, the supporters of a new paradigm seek to persuade the scientific community of its merits. Validation studies, by providing rigorous evidence of the relevance and reliability of new methods, serve as a key tool for persuading toxicologists, regulators, and industry to adopt them. Successful validation lends credibility and legitimacy to the new paradigm.
f) *Validation as a driver of methodological development:* Kuhn emphasized that as a new paradigm gains support, work proliferates to develop its methods and applications. The demand for validated new approach methodologies has spurred intense scientific activity in developing and refining *in vitro*, *in silico* and other alternative techniques. The validation imperative has been a key driver of innovation in 21st century toxicology.
g) *Validation as a facilitator of regulatory acceptance:* Formal validation is often a prerequisite for regulatory acceptance of new toxicological methods. By enabling the replacement of traditional animal tests with innovative techniques in regulatory contexts, validation helps shift the field as a whole towards the new paradigm. It provides an institutionalized mechanism for the new paradigm to infiltrate regulatory practice.
h) *Validation as a battleground for paradigms:* The standards and criteria used in validation are themselves shaped by the prevailing paradigm. Debates around validation often reflect deeper tensions between traditional and new paradigms in toxicology. How validation is conceptualized and practiced can either reinforce the *status quo* or enable more transformative change. Validation is a key arena where the paradigm shift is negotiated.

In these ways, validation is not just a technical process, but a crucial social and scientific dynamic through which the revolution in toxicology is unfolding. It provides a set of practices and institutions by which the new paradigm can establish itself and displace the old. At the same time, the nature of validation is itself contested and evolving as part of the broader paradigm shift.

4.2 Peter Singer's utilitarian approach

Peter Singer's utilitarian philosophy (Singer 1975, 2023), which emphasizes maximizing overall welfare and minimizing suffering, resonates with the validation challenge in toxicology in several ways:

a) *Reducing animal suffering:* A key tenet of utilitarianism is minimizing suffering. The use of animals in toxicity testing often involves significant suffering. By validating alternative, animal-free methods, we can reduce this suffering, which is a strong utilitarian argument for pursuing and accepting these new approaches. Validation is a necessary step to replace animal use and thereby minimize animal welfare costs.
b) *Maximizing human welfare:* For utilitarians like Singer, the ultimate goal is maximizing overall human and animal welfare. Validation of new toxicological methods can contribute to human welfare by enabling faster, cheaper and more predictive safety assessments. This can mean safer products, more efficient development of new therapies, and better protection of public health and the environment. From a utilitarian perspective, these potential welfare gains for humans must be weighed against any costs or risks of the new methods.
c) *Expanding the moral circle:* Singer's utilitarianism extends moral consideration to all sentient beings capable of suffering, including animals. This challenges us to weigh animal welfare alongside human welfare in our ethical deliberations. In the context of validation, this means the welfare of test animals cannot be discounted or ignored but must be factored into the assessment of the costs and benefits of new methods. It pushes us to actively seek validated alternatives that can reduce or replace animal use.
d) *Evidence-based assessment:* Utilitarianism is fundamentally consequentialist – it judges actions by their outcomes. This aligns with the emphasis in validation on rigorous, empirical assessment of the performance and impacts of new methods. A utilitarian approach would demand robust evidence that a new method actually reduces animal use and suffering while maintaining or enhancing human welfare via improved safety assessments. Mere intentions or assumptions are not enough – validation provides the consequentialist evidence base for adopting new methods.
e) *Challenging the status quo:* Singer's utilitarianism often leads him to question accepted practices and advocate for change when the balance of welfare considerations demands it. Similarly, the validation of new toxicological methods challenges entrenched animal testing practices and pushes for a shift to more humane and human-predictive approaches. Utilitarianism provides an ethical framework to critique the current paradigm and make the case for change.
f) *Balancing competing priorities:* Utilitarianism recognizes that there can be competing welfare considerations to weigh. In validation, there may be tensions between swift adoption of non-animal methods to reduce animal suffering versus ensuring the methods are thoroughly evaluated to protect human safety. A utilitarian approach would seek to carefully balance these short and long-term welfare implications.

While Singer's utilitarianism provides a strong ethical impetus for validation of alternative methods, it is important to note that it is not the only ethical framework relevant to the issue. Deontological principles, such as respect for animal life and the duty not to cause unnecessary harm as well as virtue-based considerations around the character of science, also come into play. Validation touches on and must navigate a range of ethical considerations beyond utilitarian calculus alone. Nonetheless, utilitarianism offers a powerful lens to highlight the moral urgency of reducing animal suffering and maximizing welfare in toxicology through robust validation of new approaches.

4.3 Hans Jonas' ethics of responsibility

Jonas' ethics of responsibility (Jonas, 1984) are based on the notion that responsibility is a more fundamental ethical principle than reciprocity. For Jonas, the archetype of responsibility is the

responsibility of parents for their children, which is non-reciprocal, as the child cannot be held responsible in the same way as the parent. Key aspects of Jonas' responsibility ethics include:

a) The imperative of responsibility arises from the vulnerability and dependence of the object of responsibility (e.g., a child) on the subject of responsibility (e.g., a parent).
b) Genuine responsibility is non-reciprocal and oriented towards the future. It is concerned with long-term consequences and the welfare of future generations.
c) In the technological age, the scope of human responsibility has expanded due to the far-reaching, cumulative, and irreversible effects of our collective actions. We have a responsibility to protect the future of humanity and the planet.
d) The "*heuristics of fear*" – we should give more weight to the threat of catastrophic consequences (e.g., environmental devastation) than to optimistic projections when making decisions about technology and progress.

In relation to validation of new toxicological methods:

– Jonas' ethics would emphasize our responsibility to ensure the safety and wellbeing of people and the environment, not just in the short term but for future generations. Validation must rigorously assess long-term consequences and err on the side of caution.
– The vulnerability of the public and ecosystems to potential harms from chemicals places a strong responsibility on regulators and scientists to ensure tests are reliable and protective. The non-reciprocal nature of this responsibility should motivate proactive and precautionary validation approaches.
– Jonas' "*heuristics of fear*" suggest giving more weight to evidence of potential hazards than to projections of safety when validating tests. Methods should be strongly challenged to prove they do not produce false negatives that could enable catastrophic consequences.
– The irreversible effects of some chemical exposures heighten responsibilities around validation. Once harm occurs, it may not be reversible, so robust pre-market validation is crucial.

In summary, Jonas' philosophy provides an ethical foundation for rigorous, public health-protective test method validation that gives more weight to long-term welfare and catastrophic threats than to short-term benefits. It places the onus on test developers to alleviate legitimate scientific doubts.

4.4 Peter Sloterdijk's critique of cynical reason

Peter Sloterdijk is a contemporary German philosopher who is difficult to categorize into a single philosophical school. Sloterdijk's *Critique of cynical reason* (1983) offers some interesting perspectives that resonate with certain critiques of the validation process for toxicological test methods:

a) *Enlightened false consciousness:* Sloterdijk argues that modern cynicism is a form of "*enlightened false consciousness*" – a situation where individuals are aware of the flaws, biases, and ideologies that shape their beliefs and actions, yet they persist in those beliefs and actions regardless. Critics might argue that some stakeholders in the validation process (e.g., industry, regulators) are aware of the limitations and value judgments inherent in animal tests yet persist in treating their results as objective truth.
b) *Institutional cynicism:* Sloterdijk discusses how cynicism can become embedded in the practices and discourses of institutions. One could argue that a degree of cynicism underlies the slow pace of adoption of new test methods – stakeholders engage in validation processes with a cynical expectation that novel methods will ultimately fail to replace entrenched animal tests, regardless of their actual performance.
c) *Contradiction between knowledge and action:* Sloterdijk highlights the disjunction between what people know and what they do. In the context of validation, there may be a contradiction between the scientific knowledge generated about the reliability and relevance of a new test (e.g., evidence that it outperforms an animal test) and the regulatory action taken (e.g., retaining the animal test as the "gold standard").
d) *Ideology masking power relations:* For Sloterdijk, cynical reason often functions to mask and perpetuate power imbalances. Critiques of validation have suggested that the process serves to entrench the position of powerful incumbent methods and stakeholders while creating barriers for novel methods under the guise of scientific rigor and objectivity.
e) *Erosion of grand narratives:* Sloterdijk posits that cynicism arises from the collapse of grand narratives and ideologies. In the context of toxicology, the historical grand narrative of animal experimentation as the bedrock of safety assessment is increasingly challenged by new scientific understandings and social values. Yet, validation processes may still be shaped by a residual cynicism that alternatives cannot fully replace traditional methods.

At the same time, it is important to recognize that Sloterdijk's critique is a broad philosophical and cultural analysis, not a specific commentary on scientific validation. The above resonances are thought-provoking parallels but should not be overextended. Ultimately, the validation process aims to serve important scientific and public health protection goals, even if it may be shaped by some of the cynical dynamics Sloterdijk discusses.

4.5 Post-modern capitalism and meta-modern philosophy

Several key philosophers and thinkers are associated with post-modern capitalism and metamodern philosophy[10]. These thinkers have contributed to understanding the complexities of late capi-

[10] *Postmodern capitalism:* Post-modern capitalism: Jean-François Lyotard – criticized grand narratives and emphasized the role of language games in society; Fredric Jameson – analyzed late capitalism and its cultural logic. Jean Baudrillard – explored concepts of hyperreality and simulation in consumer society. Gilles Deleuze and Félix Guattari – developed ideas about deterritorialization and rhizomatic structures in capitalism. Zygmunt Bauman – wrote about liquid modernity and the shifting nature of social structures. David Harvey – analyzed the condition of postmodernity and flexible accumulation.

Metamodern philosophy: Timotheus Vermeulen and Robin van den Akker – coined the term "metamodernism" and outlined its key principles. Luke Turner – developed the *Metamodernist Manifesto*. Hanzi Freinacht (pseudonym for Daniel Görtz and Emil Ejner Friis) – wrote about metamodern political theory and cultural development. Seth Abramson – explored metamodernism in literature and culture. Alexandra Dumitrescu – analyzed metamodernism in relation to spirituality and consciousness. Brent Cooper – developed ideas about metamodern systems theory and social change.

talism and the emerging cultural and philosophical responses to postmodernism, shaping our understanding of contemporary society and thought.

a) *Globalization:* The global spread of capitalism has led to the integration of economies, cultures, and political systems around the world. This has facilitated the movement of capital, goods, services, and labor across borders, creating a more interconnected global economy. Globalization has also resulted in the rise of multinational corporations and the global division of labor. We discussed the impact of globalization on alternative methods in Bottini et al. (2007).
b) *Technological advancements:* The digital economy – the rapid pace of technological innovation, especially in information and communication technologies, has transformed the economic landscape. In particular, AI has emerged as a powerful new tool. We have discussed this in the context of NAMs in Hartung (2023a,b) and Kleinstreuer and Hartung (2024).
c) *Consumer culture and media saturation:* The rise of consumer culture, where identity and social status are significantly influenced by consumer choices. Media play a crucial role in shaping desires and norms, leading to a saturation of cultural life with (commercial) messages. We touched on this in von Aulock et al. (2022).
d) *Financialization:* The increasing importance of financial markets, financial motives, financial institutions, and financial elites in the operation of the economy and its governing institutions. This was discussed in Bottini and Hartung (2009) and Meigs et al. (2018).

The resonance of other themes with our field still needs to be explored: *Flexibilization and precarity* (the increasing flexibility of the labor market with implications for job security and social stability), *de-industrialization and the service economy* (the shift towards services and knowledge-based industries), and *cultural and ideological shifts* (the skepticism towards grand narratives and ideologies). Some of these concepts have been transformed to what is now called meta-modern philosophy. Key features of meta-modern philosophy include *oscillation* between modernist and postmodernist attitudes, acknowledging the value and limitations of both. This oscillation is not a balance but a dynamic movement that allows for embracing sincerity and irony, optimism and skepticism, engagement and detachment. Metamodernism reintroduces *sincerity*, *hope*, *and earnestness* into cultural and philosophical discourse. *Constructive engagement* seeks to move beyond critique and deconstruction towards constructive engagement. It encourages the pursuit of new meanings, values, and narratives that can address contemporary challenges, even while recognizing the fluid and constructed nature of these narratives. *Informed naivety* adopts a stance of a deliberate choice to engage with the world in a hopeful and optimistic manner, fully aware of the potential for disillusionment and failure. *Embracing complexity* recognizes the interconnected and interdependent nature of global challenges in a way that is both critical and hopeful. *Hybridization* blends genres, styles, and disciplines, reflecting the interconnected and multifaceted nature of contemporary life.

5 Ethical principles relevant to the validation of testing methods

The author is involved in an ongoing discussion on the ethics of regulation of chemicals, expanding on Bhuller et al. (2024, submitted). From these discussions the ethical principles listed in Box 2 emerged.

Box 2: Key ethical principles relevant to the regulation of drugs and chemicals

1. *Beneficence:* The obligation to promote the well-being of individuals and society. Regulatory decisions should prioritize maximizing benefits and minimizing harms associated with drugs and chemicals.
2. *Nonmaleficence:* Linked closely to beneficence, nonmaleficence underscores the imperative to avoid causing harm.
3. *Autonomy:* Respect for autonomy entails recognizing individuals' right to make informed decisions about their health and well-being. Regulatory frameworks should uphold the autonomy of consumers by providing accurate information and ensuring transparency in decision-making processes.
4. *Justice:* The principle of justice calls for fair and equitable distribution of benefits and burdens. Regulatory agencies must strive to ensure that access to safe and effective drugs and chemicals is not disproportionately skewed based on factors such as socioeconomic status, geography, or ethnicity.
5. *Informed consent:* Informed consent is essential in clinical research and medical practice. Regulatory requirements often mandate that individuals are adequately informed about the risks and benefits of drugs and chemicals before consenting to their use, particularly in the context of clinical trials.
6. *Privacy and confidentiality:* Protecting individuals' privacy and confidentiality is crucial in regulatory activities involving sensitive health information.
7. *Transparency and accountability:* Transparency in regulatory processes fosters trust and accountability. Regulatory agencies should operate in an open and transparent manner, disclosing relevant information to the public and stakeholders and soliciting input from diverse perspectives.
8. *Integrity and honesty:* Upholding integrity and honesty is fundamental to maintaining public trust in regulatory systems.

Tailoring these and the evidence-based paradigm to validation in the context here is summarized in Figure 2.

The ethical validation of NAMs is underpinned by several key principles derived from the 3Rs framework of replacing, reducing, and refining animal use (Balls et al., 2024) and broader bioethical guidelines:

a) *Respect for animal welfare:* The principle of respect for animal welfare asserts that animals have intrinsic value and the capacity to suffer, giving rise to a moral imperative to replace, reduce, and refine animal use in testing. In the context of NAM validation, this translates into prioritizing methods that can fully replace animal tests where scientifically feasible. Where full replacement is not yet possible, NAMs that can reduce the number of animals needed or refine procedures to minimize pain and distress should be pursued. The ultimate goal is to shift testing approaches away from animal use in favor of more human-relevant and humane methods.
b) *Scientific validity and fitness-for-purpose:* The ethical imperative to replace animal testing with NAMs can only be fulfilled if the alternative methods are scientifically valid for their intended purpose. Rigorous validation is required to establish that a NAM is a reliable and relevant replacement for the animal test, providing equivalent or better information to inform safety decisions. Rather than taking a check-box approach, validation should flexibly consider the specific context of use and adopt a "fit-for-purpose" mentality. This involves qualifying NAMs for specific applications and incorporating them into integrated testing strategies, rather than demanding 1:1 replacement of an animal test.
c) *Proportionality of validation requirements:* The principle of proportionality asserts that the level of validation required for a NAM should be commensurate with its expected benefits and limitations. Less burdensome validation may be appropriate for NAMs intended for screening or prioritization compared to methods used for regulatory safety assessments. Validation bodies and resulting test guidance should avoid imposing requirements so rigorous that developing NAMs becomes prohibitive; some space should be left for exploration and iterative learning with new technologies in constrained domains. At the same time, scientific rigor cannot be unduly relaxed without endangering human safety. Maintaining the right balance demands judgement but is critical to the ethical-scientific purpose of NAM validation.
d) *Precaution regarding uncertainties:* The precautionary principle advocates erring on the side of caution when there is uncertainty about potential risks. In NAM validation, this supports a stepwise progression with independent review at each stage to identify unanswered questions. Where a NAM relies on a novel technology or biological mechanism, more evidence may initially be required to confirm relevance for humans compared to a method that closely mimics a well-characterized animal test. Until the NAM is definitively proven valid for a given purpose, it may be appropriate to retain some traditional testing as a backstop to avoid overlooking potential hazards.
e) *Transparency and public participation:* Ethical validation of NAMs requires an open, accountable process with input from all relevant stakeholders. Transparency about validation study design, criteria, results and conclusions enables independent verification and supports scientific credibility. Allowing public comment periods and engaging with affected communities ensures that societal perspectives on the acceptability of technical choices are considered. Animal welfare advocates in particular should have a role in NAM validation decisions. Overall, transparency and participation build trust that validation is an objective process reflecting both scientific and ethical priorities.
f) *Commitment to continuous improvement:* Ethical principles demand a continuing commitment to improve NAMs even after initial validation, as new knowledge and technology emerges. Post-validation monitoring of real-world performance, coupled with an openness to refining NAMs based on user feedback and evolving scientific understanding, ensures optimal benefits over time. An important focus should be expanding the applicability domain of NAMs to enable more animal testing to be replaced, where initial validation focused on narrow uses. Dedicating resources to NAM improvement over the lifecycle underscores that validation is not an endpoint but rather a key milestone on a longer journey.

Mancini and Nannoni (2022) reinforce the argument that traditional frameworks like the 3Rs are no longer sufficient in addressing contemporary ethical concerns in animal research. They suggest that new ethical principles and frameworks, such as animal-centered ethics frameworks, should be integrated into the discussion of animal research. This argues for a more holistic approach to ethical testing that respects the autonomy and welfare of animal participants while ensuring scientific integrity. The paper, while not addressing validation, provides a valuable perspective by emphasizing the importance of considering animal consent and welfare as central components of ethical validation frameworks.

The 12Rs Framework proposed by Brink and Lewis (2023) offers a comprehensive expansion of the traditional 3Rs, integrating principles of animal welfare, scientific integrity, and social values into a unifying ethical construct. The 12Rs are grouped into three main categories: Animal welfare Rs (AWRs), social value Rs (SVRs), and scientific integrity Rs (SIRs), which intersect to form additional principles such as righteousness, reliability, and reckoning. This structured approach parallels the need for ethical frameworks in the context of the validation and adoption of NAMs as alternatives to animal testing. This aligns with the current discussion on the philosophical underpinnings of validation, as it highlights the necessity for a holistic approach to evaluating and implementing NAMs. The framework can serve as a model for ensuring that validation processes are not only scientifically robust but also ethically and socially responsible, thus addressing broader ethical considerations that are crucial for regulatory acceptance and public trust in NAMs. The framework acts as a mind map to help navigate the complexities of animal research ethics. This aligns with our focus on providing a systematic approach to validation processes that is both scientifically robust and ethically sound.

In summary, validation of NAMs as alternatives to animal testing must balance several ethical principles and competing priorities. A progressive, flexible, context-specific yet scientifically robust approach to establishing fit-for-purpose validity is needed. The ultimate guidepost should be reducing animal suffering while protecting human safety through scientifically credible and human-relevant methods. Engaging diverse stakeholders in an open

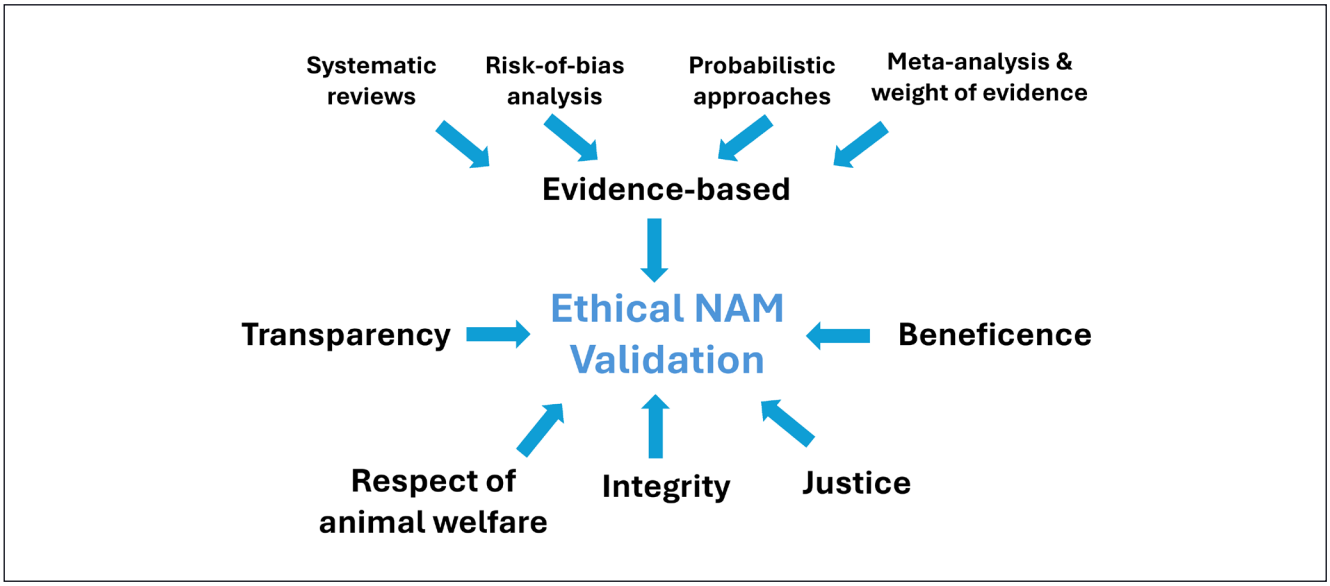

Fig. 2: Aspects of an evidence-based ethical framework for next generation validation

validation process with shared commitment to continuous improvement offers the best path to achieve this vision.

The core ethical principles guiding the validation of NAMs as alternatives to animal testing include beneficence, non-maleficence, justice, and respect for animal welfare. Beneficence refers to maximizing the benefits of NAMs while minimizing potential harms through rigorous testing, comprehensive evaluation, and responsible oversight even after validation[11]. Non-maleficence or "do no harm" entails protecting against foreseeable risks to human health and the environment by ensuring NAMs provide reliable and relevant information for safety assessments (Hartung and Tsatsakis, 2021). Justice involves ensuring the fair and equitable development and application of NAMs, preventing disproportionate impacts on vulnerable populations, and considering the broader societal implications of transitioning away from animal testing[12]. Respect for animal welfare means recognizing the intrinsic value of animals and their capacity to suffer, and therefore prioritizing the development and use of NAMs that replace, reduce, or refine animal use in testing (Bayne and Morris, 2012). Additional relevant principles are proportionality of validation requirements to the expected benefits and limitations of the NAM, accountability of validation bodies and method developers, transparency about the validation process and results, integrity and honesty in reporting, and precaution when there is uncertainty about the reliability or relevance of a NAM for a specific context of use (NASEM, 2023).

Navigating tensions around scientific rigor versus speed of implementation, individual NAM validity versus applicability to testing strategies, and traditional approaches versus new technologies is an inherent challenge. Addressing these tensions requires open dialogue, adaptability, and a focus on upholding both human and animal wellbeing as the ultimate priorities[12]. Overall, these shared ethical principles aim to guide the complex process of establishing scientific confidence in NAMs as alternatives to animal testing (NASEM, 2023).

6 Conclusions and the way forward

The ongoing revision of the OECD validation process has stimulated discussion. Gourmelon et al. (2024) give a positive outlook:

- Increased awareness of the need for test method validation among stakeholders.
- Various stakeholders are organizing themselves to coordinate validation activities, including private entities, public-private partnerships, and publicly funded research programs.
- Growing acceptance that all stakeholders should play a role and participate in funding.
- Recognition of the high level of expertise required in managing validation activities.
- Discussions on optimizing the ring-trial process to balance costs and benefits.
- Acknowledgment that the benefits of having validated methods allowing mutual acceptance of data outweigh the costs.
- Estimated annual savings from the Mutual Acceptance of Data system are around 309 million EUR, demonstrating that validation is a profitable investment.
- Proposals for more efficient validation processes, such as giving more weight to well-conducted and documented transferability studies.
- Suggestions for cost reduction, including centralized repositories for reference chemicals and streamlined data management.

[11] https://nap.nationalacademies.org/catalog/26496/new-approach-methods-nams-for-human-health-risk-assessment-proceedings

[12] https://issues.org/ethics-in-animal-research-perspective-johnson-forum/

– Increased focus on the readiness of methods before entering validation, potentially reducing time and resources needed.

The paper concludes that while challenges remain, there is a positive trend towards shared responsibility and recognition of the value of validation in the development and acceptance of new approach methods for toxicity testing.

The discussions in our article show that the current validation paradigm, based on the 3R principle (replacement, reduction, and refinement), which has been the cornerstone of animal research ethics for decades, is now facing substantial limitations in the context of evolving scientific, ethical, and societal expectations. The integration of concepts such as the 12Rs Framework by Brink and Lewis (2023), the animal-centered research ethics proposed by Mancini and Nannoni (2022), and the introduction of a scientific confidence framework reveals a pressing need to fundamentally rethink and revise the validation process. Key points include:

a) *Expanding beyond the traditional 3Rs:* The 12Rs Framework and the principles of animal-centered ethics illustrate the necessity to extend ethical considerations beyond mere welfare, emphasizing animal autonomy, societal values, and scientific integrity. These frameworks call for a broader, more holistic approach to animal research ethics that better aligns with contemporary views on animal welfare and societal expectations. This expansion necessitates a corresponding shift in the validation process to accommodate these broader ethical and social considerations.
b) *Embracing flexibility and context-specific validation:* The validation process, as traditionally conceived, has been criticized for its rigidity and reliance on animal-based models as the gold standard. The shift towards a fit-for-purpose validation approach emphasizes the need for context-specific evaluation, recognizing that different NAMs may require tailored validation criteria depending on their intended application. This approach supports a more adaptive and iterative process, making validation a more dynamic tool that evolves in tandem with scientific and technological advancements.
c) *Integration of scientific confidence and transparency:* The concept of a scientific confidence framework extends validation by incorporating aspects such as reliability, relevance, and transparency. It emphasizes the need for rigorous yet flexible validation criteria that support the scientific credibility of NAMs while accommodating emerging methodologies. Transparency, including open communication of validation processes and outcomes, fosters greater trust among stakeholders and encourages more widespread acceptance of NAMs.
d) *Harmonizing ethical and scientific goals:* The discussion highlights a key tension between ethical imperatives – such as reducing animal suffering – and the scientific goals of ensuring human safety and environmental protection. Harmonizing these goals requires a redefined validation process that not only establishes scientific robustness but also upholds ethical standards. This alignment could be achieved through frameworks that incorporate principles from both traditional ethics (e.g., utilitarianism and Kantian ethics) and contemporary frameworks like the animal-centered ethics proposed by Mancini and Nannoni (2022).
e) *Revising the validation paradigm:* A revised validation paradigm should incorporate elements from evidence-based toxicology, systematic reviews, and meta-analyses, as well as structured WoE approaches to integrate diverse data sources. This would create a more comprehensive and robust validation framework that better supports regulatory acceptance and public trust. Additionally, the paradigm should allow for post-validation monitoring and continuous improvement to keep pace with scientific and technological progress.

If we take these aspects together with the two accompanying papers (Hartung et al., 2024a,b), which address how to make validation more human-relevant by biomarkers and more efficient using AI, Figure 3 emerges.

Validation emerges as a somewhat fluid concept that can adapt the needs of a given study to a core of unchangeable values and principles. This flexible response will require new levels of transparency and versioning of tests and their validity status. We might need to discuss a concept of "probability of validity". And we will need to define what "context of use" really means if we adapt the concept from the realm of drug development. This could include:

– Severeness of the hazard/disease outcome
– Type of application of substances and resulting exposures including likely co-exposures
– Applicability to mixtures
– Relevance of high dose (e.g., work safety)
– Consideration of vulnerable sub-populations

All these and probably many more aspects could factor into the expectations to be set for qualifying as a test for a given purpose.

Our paper focuses on ethical and philosophical aspects, with less attention given to technical or regulatory challenges in validation, which we have discussed before. By connecting validation practices to broader philosophical traditions like utilitarianism, ethics of responsibility, and post-modern thought, the paper aims to elevate the discussion beyond technical considerations to examine fundamental assumptions and values. Using Kuhn's model of scientific revolutions, the paper (again) frames the move towards NAMs as a paradigm shift, helping to explain resistance to change and the transformative potential of new methods. There are a number of limitations to this discussion by a non-philosopher: The paper is largely theoretical and could benefit from more concrete examples or case studies illustrating how philosophical principles translate into practical validation decisions. While it covers a wide range of philosophical perspectives, the connections to validation are sometimes tenuous or speculative.

In conclusion, the convergence of new ethical frameworks and scientific advancements necessitates a reformation of the traditional validation process. By integrating principles such as the 12Rs Framework, animal-centered ethics, and scientific confidence frameworks, the validation process can evolve into a more flexible, context-specific, and ethically sound practice. This evolution will better support the acceptance and implementation of NAMs, ultimately leading to a more humane and scientifically credible approach to toxicological testing that prioritizes both human and animal welfare. We point towards potential evolutions in validation practices, such as greater incorporation of evidence-based toxicology principles and more adaptive, iterative processes. Such a reim-

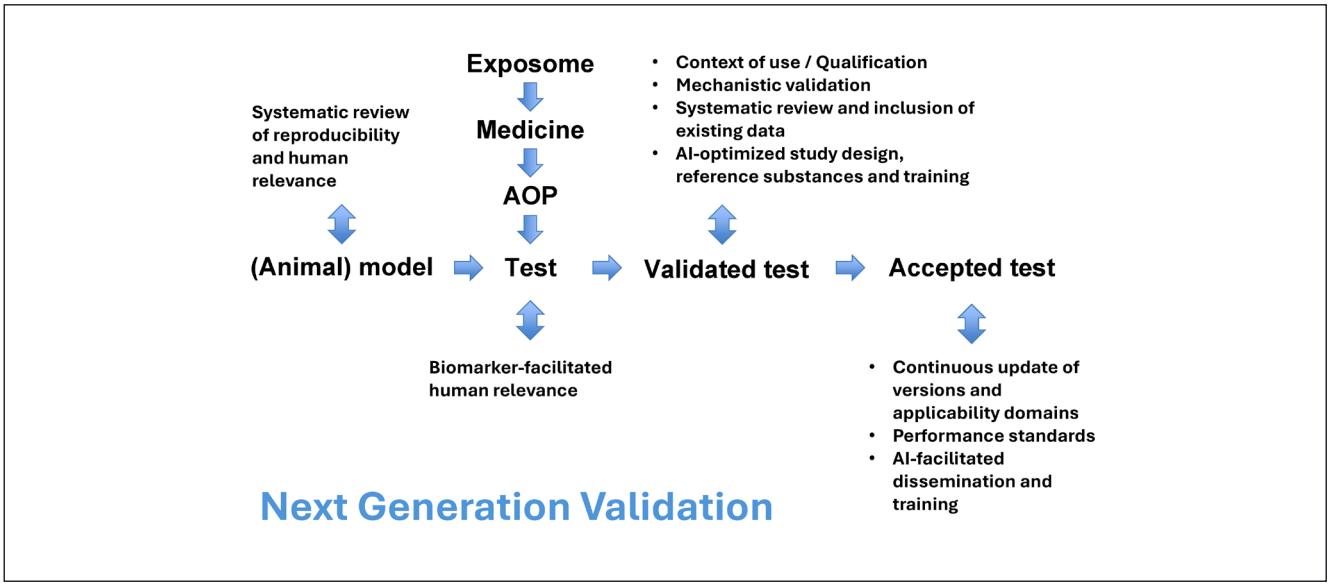

Fig. 3: Next generation validation

The diagram integrates concepts from this paper, Hartung et al. (2004, 2009, 2013a, 2024a,b), and Sillé et al. (2020, 2024). It starts with a call to systematically assess the performance of the model that is to be replaced. It also considers that needs for tests might originate from identified adverse outcome pathways (AOPs), i.e., an exposure-to-disease hypothesis in medicine, which could be fueled by the exposome as more mechanistic understanding of the exposure side of disease. By embracing the biomarker concept, more clinically relevant endpoints are favored. The validation process itself can borrow the clear context of use definition from the biomarker concept; this might change over time and require performance standards and version-tracking once accepted. Corresponding to the more pathway- and clinical outcome-based biomarker approach, a mechanistic validation focusing on demonstrating that the right AOPs are reflected is suggested. The use of existing data, ideally retrieved by systematic review, is encouraged. AI-facilitated validation study design and execution (e-validation) shall make this process more efficient. Once accepted, it needs to be understood that with additional contexts of use or learnings about restrictions (changing applicability domain) the use of a test might change ("incremental validation"); furthermore, variants of the test and new tests based on the same test principle might emerge; this requires version tracking as a new role for validation bodies.

agined validation framework not only addresses current challenges but also positions the field to navigate future advancements and ethical considerations with greater agility and responsiveness.

References

Balls, M., Blaauboer, B., Brusick, D. et al. (1990). Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures. *Altern Lab Anim 18*, 313-337. doi:10.1177/026119299001800131.1

Balls, M., Blaauboer, B. J., Fentem, J. H. et al. (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations of ECVAM workshop 5. *Altern Lab Anim 23*, 129-147. doi:10.1177/026119299502300116

Balls, M., Bass, R., Curren, R. et al. (2024). 60 Years of the 3Rs symposium: Lessons learned and the road ahead. *ALTEX 41*, 179-201. doi:10.14573/altex.2403061

Bayne, K. and Morris, T. (2012). Laws, regulations and policies relating to the care and use of nonhuman primates in biomedical research. In C. R. Abee, K. Mansfield, S. Tardif et al. (eds), *Nonhuman Primates in Biomedical Research: Biology and Management*. Academic Press. doi:10.1016/B978-0-12-381365-7.00002-9

Bhuller, Y., Deonandan, R. and Krewski, D. (2024). Relevance and feasibility of principles for health and environmental risk decision-making. *J Toxicol Environ Health B 27*, 189-211. doi:10.1080/10937404.2024.2338078

Bhuller, Y., Hilton, G., Avey, M. et al. (submitted). Ethical principles for regulatory risk decision-making.

Bottini, A. A., Amcoff, P. and Hartung, T. (2007). Food for thought ... on globalisation of alternative methods. *ALTEX 24*, 255-269. doi:10.14573/altex.2007.4.255

Bottini, A. A., Alepee, N., De Silva, O. et al. (2008). Optimization of the post-validation process. The report and recommendations of ECVAM workshop 67. *Altern Lab Anim 36*, 353-366.

Bottini, A. A. and Hartung, T. (2009). Food for thought... on the economics of animal testing. *ALTEX 26*, 3-16. doi:10.14573/altex.2009.1.3

Bouhifd, M., Andersen, M. E., Baghdikian, C. et al. (2015). The human toxome project. *ALTEX 32*, 112-124. doi:10.14573/altex.1502091

Bower, J. F., McClung, J. B., Watson, C. et al. (2014). Recommendations and best practices for reference standards and reagents used in bioanalytical method validation. *AAPS J 16*, 352-356. doi:10.1208/s12248-014-9566-y

Brink, C. B. and Lewis, D. I. (2023). The 12 Rs framework as a

comprehensive, unifying construct for principles guiding animal research ethics. *Animals 13*, 1128. doi:10.3390/ani13071128

Browne, P., Delrue, N. and Gourmelon, A. (2019). Regulatory use and acceptance of alternative methods for chemical hazard identification. *Curr Opin Toxicol 15*, 18-25. doi:10.1016/j.cotox.2019.02.003

Caloni, F., De Angelis, I. and Hartung, T. (2022). Replacement of animal testing by integrated approaches to testing and assessment (IATA): A call for in vivitrosi. *Arch Toxicol 96*, 1935-1950. doi:10.1007/s00204-022-03299-x

Clippinger, A. J., Raabe, H. A., Allen, D. G. et al. (2021). Human-relevant approaches to assess eye corrosion/irritation potential of agrochemical formulations. *Cutan Ocul Toxicol 40*, 145-167. doi:10.1080/15569527.2021.1910291

Frazier, J. M. (1994). The role of mechanistic toxicology in test method validation. *Toxicol In Vitro 8*, 787-791. doi:10.1016/0887-2333(94)90068-X

Goldberg, A.M. (1986). *In Vitro Toxicology: Approaches to Validation*. New York, NY, USA: Mary Ann Liebert, Inc.

Goldberg, A. M., Frazier, J. M., Brusick, D. et al. (1993). Framework for validation and implementation of in vitro toxicity tests. *In Vitro Cell Dev Biol Anim 29*, 688-692. doi:10.1007/bf02631424

Golden, E., Allen, D. and Amberg, A. (2024). Toward implementing virtual control groups in nonclinical safety studies: Workshop report and roadmap to implementation. *ALTEX 41*, 282-301. doi:10.14573/altex.2310041

Gourmelon, A., Hubert, P., Grignard, E. et al. (2024). The benefits of validation of methods for toxicity testing outweigh its costs. *ALTEX 41*, 395-401. doi:10.14573/altex.2403051

Grinnel, F. (2009). *Everyday Practice of Science: Where Intuition and Passion Meet Objectivity and Logic*. Oxford University Press. doi:10.1093/acprof:oso/9780195064575.001.0001

Hartung, T. and Spielmann, H. (1995). Der lange Weg zur validierten Ersatzmethode (The sophisticated process of validation) [Article in German]. *ALTEX 12*, 98-103. https://www.altex.org/index.php/altex/article/view/1677

Hartung, T., Bremer, S., Casati, S. et al. (2004). A modular approach to the ECVAM principles on test validity. *Altern Lab Anim 32*, 467-472. doi:10.1177/026119290403200503

Hartung, T. (2007). Food for thought ... on validation. *ALTEX 24*, 67-80. doi:10.14573/altex.2007.2.67

Hartung T. (2008). Towards a new toxicology – Evolution or revolution? *Altern Lab Anim 36*, 635-639. doi:10.1177/026119290803600607

Hartung T. (2009). Food for thought … on evidence-based toxicology. *ALTEX 26*, 75-82. doi:10.14573/altex.2009.2.75

Hartung, T. (2010). Evidence-based toxicology – The toolbox of validation for the 21st century? *ALTEX 27*, 253-263. doi:10.14573/altex.2010.4.253

Hartung, T., Stephens, M. and Hoffmann, S. (2013a). Mechanistic validation. *ALTEX 30*, 119-130. doi:10.14573/altex.2013.2.119

Hartung, T., Luechtefeld, T., Maertens, A. and Kleensang, A. (2013b). Integrated testing strategies for safety assessments. *ALTEX 30*, 3-18. doi:10.14573/altex.2013.1.003

Hartung, T. and Tsatsakis, A. M. (2021). The state of the scientific revolution in toxicology. *ALTEX 38*, 379-386. doi:10.14573/altex.2106101

Hartung, T. and Krewski, D. (2022). Development of an evidence-based risk assessment framework. *ALTEX 39*, 442. doi:10.14573/altex.22S2

Hartung T. (2023a). Artificial intelligence as the new frontier in chemical risk assessment. *Front Artif Intell 6*, 1269932. doi:10.3389/frai.2023.1269932

Hartung T. (2023b). ToxAIcology – The evolving role of artificial intelligence in advancing toxicology and modernizing regulatory science. *ALTEX 40*, 559-570. doi:10.14573/altex.2309191

Hartung, T. (2023c). Evidence-based toxicology. In *Encyclopedia of Toxicology* (4th edition, 561-565). doi:10.1016/B978-0-12-824315-2.00973-8

Hartung, T. and Tsaioun, K. (2024). Evidence-based approaches in toxicology: Their origins, challenges, and future directions. *Evid Based Toxicol*, in press.

Hartung, T., Maertens, A. and Luechtefeld, T. (2024a). E-validation – Unleashing AI for validation. *ALTEX 41*, 567-587. doi:10.14573/altex.2409211

Hartung, T., King, N. P. M., Kleinstreuer, N. et al. (2024b). Leveraging biomarkers and translational medicine for preclinical safety - Lessons for advancing the validation of alternatives to animal testing. *ALTEX 41*, 545-566. doi:10.14573/altex.2410011

Hoffmann, S. and Hartung, T. (2006a). Designing validation studies more efficiently according to the modular approach: Retrospective analysis of the EPISKIN test for skin corrosion. *Altern Lab Anim 34*, 177-191. doi:10.1177/026119290603400209

Hoffmann, S. and Hartung, T. (2006b). Towards an evidence-based toxicology. *Human Exp Toxicol 25*, 497-513. doi:10.1191/0960327106het648oa

Hoffmann, S., Edler, L., Gardner, I. et al. (2008). Points of reference in the validation process. *Altern Lab Anim 36*, 343-352. doi:10.1177/026119290803600311

Hoffmann, S., de Vries, R. B. M., Stephens, M. L. et al. (2017). A primer on systematic reviews in toxicology. *Arch Toxicol 91*, 2551-2575. doi:10.1007/s00204-017-1980-3

ICCVAM – Interagency Coordinating Committee on the Validation of Alternative Methods (2018). A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States. doi:10.22427/ntp-iccvam-roadmap2018

Jonas, H. (1984). *The Imperative of Responsibility in Search of an Ethics for the Technological Age*. Chicago, London: The University of Chicago Press.

Judson, R., Kavlock, R., Martin, M. et al. (2013). Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *ALTEX 30*, 51-56. doi:10.14573/altex.2013.1.051

Kleensang, A., Maertens, A., Rosenberg, M. et al. (2014). Pathways of Toxicity. *ALTEX 31*, 53-61. doi:10.14573/altex.1309261

Kleinstreuer, N. and Hartung, T. (2024). Artificial intelligence (AI) – It's the end of the tox as we know it (and I feel fine). *Arch Toxicol 98*, 735-754. doi:10.1007/s00204-023-03666-2

Krewski, D., Saunders-Hastings, P., Baan, R. A. et al. (2022). Development of an evidence-based risk assessment framework. *ALTEX 39*, 667-693. doi:10.14573/altex.2004041

Leist, M., Hasiwa, N., Daneshian, M. et al. (2012). Validation and quality control of replacement alternatives – Current status and future challenges. *Toxicol Res 1*, 8-22. doi:10.1039/c2tx20011b

Leist, M. and Hartung, T. (2013). Inflammatory findings on species extrapolations: Humans are definitely no 70-kg mice. *Arch Toxicol 87*, 563-567. doi:10.1007/s00204-013-1038-0

Linkov, I., Massey, O., Keisler, J. et al. (2015). From "weight of evidence" to quantitative data integration using multicriteria decision analysis and Bayesian methods. *ALTEX 32*, 3-8. doi:10.14573/altex.1412231

Maertens, A., Golden, E., Luechtefeld, T. et al. (2022). Probabilistic risk assessment – The keystone for the future of toxicology. *ALTEX 39*, 3-29. doi:10.14573/altex.2201081

Maertens, A., Antignac, E., Benfenati, E. et al. (2024a). The probable future of toxicology – Probabilistic risk assessment. *ALTEX 41*, 273-281. doi:10.14573/altex.2310301

Mancini, C. and Nannoni, E. (2022). Relevance, impartiality, welfare and consent: Principles of an animal-centered research ethics. *Front Anim Sci 3*, 800186. doi:10.3389/fanim.2022.800186

Meigs, L., Smirnova, L., Rovida, C. et al. (2018). Animal testing and its alternatives – The most important omics is economics. *ALTEX 35*, 275-305. doi:10.14573/altex.1807041

Mondou, M., Maguire, S., Pain, G. et al. (2021). Envisioning an international validation process for new approach methodologies in chemical hazard and risk assessment. *Environ Adv 4*, 100061. doi:10.1016/j.envadv.2021.100061

NASEM – National Academies of Sciences, Engineering, and Medicine (2023). *Building Confidence in New Evidence Streams for Human Health Risk Assessment: Lessons Learned from Laboratory Mammalian Toxicity Tests*. Washington, DC: The National Academies Press. (Especially Chapter 5: Issues in developing a scientific confidence framework for NAM). doi:10.17226/26906

OECD (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. *Series on Testing and Assessment, No. 34*. OECD Publishing, Paris.

Parish, S. T., Aschner, M., Casey, W. et al. (2020). An evaluation framework for new approach methodologies (NAMs) for human health safety assessment. *Regul Toxicol Pharmacol 112*, 104592. doi:10.1016/j.yrtph.2020.104592

Patterson, E. A., Whelan, M. P. and Worth, A. P. (2021). The role of validation in establishing the scientific credibility of predictive toxicology approaches intended for regulatory application. *Comput Toxicol 17*, 100144. doi:10.1016/j.comtox.2020.100144

Rovida, C., Alépée, N., Api, A. M. et al. (2015). Integrated testing strategies (ITS) for safety assessment. *ALTEX 32*, 171-181. doi:10.14573/altex.1506201

Sillé, F. C. M., Karakitsios, S., Kleensang, A. et al. (2020). The exposome – A new approach for risk assessment. *ALTEX 37*, 3-23. doi:10.14573/altex.2001051

Sillé, F. C. M., Busquet, F., Fitzpatrick, S. et al. (2024). The implementation moonshot project for alternative chemical testing (IMPACT) toward a human exposome project. *ALTEX 41*, 344-362. doi:10.14573/altex.2407081

Singer, P. (1975). *Animal Liberation*. London, UK: The Bodley Head.

Singer, P. (2023). *Animal Liberation Now*. Bodley Head Childrens.

Sloterdijk, P. (1983). *Critique of Cynical Reason*. Minneapolis, USA: University Of Minnesota Press.

Spielmann, H., Liebsch, M. and Reinhardt, C. (1998). ERGATT/ECVAM workshop on acceptance of validated alternative methods: Amden III [Article in German]. *ALTEX 15*, 18-22. https://altex.org/index.php/altex/article/view/1571

Steger-Hartmann, T., Kreuchwig, A., Vaas, L. et al. (2020). Introducing the concept of virtual control groups into preclinical toxicology testing. *ALTEX 37*, 343-349. doi:10.14573/altex.2001311

Stephens, M. L., Andersen, M., Becker, R. A. et al. (2013). Evidence-based toxicology for the 21st century: Opportunities and challenges. *ALTEX 30*, 74-104. doi:10.14573/altex.2013.1.074

Stokes, W. S. and Schechtman, L. M. (2007). Validation and regulatory acceptance of new, revised, and alternative toxicological methods. In A. W. Hayes (ed.), *Principles and Methods of Toxicology* (1103-1128). Philadelphia, Pennsylvania, USA: Taylor and Francis.

van der Zalm, A. J., Barroso, J., Browne, P. et al. (2022). A framework for establishing scientific confidence in new approach methodologies. *Arch Toxicol 96*, 2865-2879. doi:10.1007/s00204-022-03365-4

von Aulock, S., Busquet, F., Locke, P. et al. (2022). Engagement of scientists with the public and policymakers to promote alternative methods. *ALTEX 39*, 543-559. doi:10.14573/altex.2209261

Worth, A. P. and Balls, M. (2004). The principles of validation and the ECVAM validation process. *Altern Lab Anim 32*, 623-629. doi:10.1177/026119290403201s105

Conflict of interest

The author declares no conflict of interest.

Data availability

No novel data was produced for this manuscript.

Acknowledgements

Thirty years of discussions with friends and colleagues on validation are gratefully appreciated. Funding was received from the European Union's Horizon 2020 research and innovation program under grant agreements No. 963845 (ONTOX) and No. 101057014 (PARC), and from the IHI-2 project Vict3R.