"The scientist is not a person who gives the right answers, he's one who asks the right questions"
Claude Lévi-Strauss, French anthropologist (1908-2009)

"If you always do what you always did, you will always get what you always got."
Attributed to different individuals[1]

Food for Thought ...

Leveraging Biomarkers and Translational Medicine for Preclinical Safety – Lessons for Advancing the Validation of Alternatives to Animal Testing

Thomas Hartung[1,2], Nicholas M. P. King[3], Nicole Kleinstreuer[4], Marcel Leist[2,5] and Danilo A. Tagle[6]

[1]Center for Alternatives to Animal Testing (CAAT), Doerenkamp-Zbinden-Chair for Evidence-based Toxicology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; [2]CAAT-Europe, University of Konstanz, Germany; [3]Critical Path Institute, Tucson, AZ, USA; [4]National Institute of Health, National Institute for Environmental Health Sciences, DTT/NICEATM, Durham, NC, USA; [5]In vitro Toxicology and Biomedicine, University of Konstanz, Konstanz, Germany; [6]National Institute of Health, National Center for Advancing Translational Sciences (NCATS), Bethesda, MD, USA

Received October 1, 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)

ALTEX 41(4), 545-566.
doi:10.14573/altex.2410011

Abstract

This article explores the potential of principles established in translational medicine for the use of biomarkers to advance the validation of alternatives to animal testing in preclinical safety assessment. It examines especially how such principles can enhance the predictive power, mechanistic understanding, and human relevance of new approach methodologies (NAMs). Key concepts from translational medicine, such as fit-for-purpose validation, evidence-based approaches, and integrated testing strategies, are already being applied to the development and validation of NAMs. The article discusses challenges in implementing biomarker-based approaches, including standardization, demonstration of relevance, regulatory acceptance, and addressing biological complexity. It also highlights opportunities for advancement through collaborative efforts, technological innovations, and regulatory evolution. Case studies demonstrate successful applications of biomarkers in preclinical safety, while future perspectives explore emerging trends like multi-omics integration, microphysiological systems, and artificial intelligence. The article emphasizes the potential of biomarkers and translational science approaches in creating more predictive, efficient, and ethical preclinical safety assessment paradigms in the use of NAMs. Use of biomarkers can enable the mechanistic validation of human-relevant models and provide a means to relate changes in NAMs to animal or clinical study results. By leveraging these tools, the field can work towards reducing reliance on animal testing while improving the accuracy and human relevance of safety predictions.

Plain language summary

This article examines how biomarkers and translational science principles can improve safety testing without using animals. Biomarkers are quantifiable indicators of biological processes. Some of these can predict disease progression or drug effects. Translational science aims to apply laboratory findings towards clinical benefits. The article explores how combining these approaches can create better, more human-relevant and validated alternatives to animal testing. It discusses challenges that the field faces, including standardization of methods and getting regulatory acceptance. It also highlights opportunities, like integration with emerging technologies and increased global collaboration. The ultimate goal is to improve human health by streamlining NAM validation processes, i.e., show that new safety tests are more accurate, efficient, and ethical than current animal-based methods.

[1] https://quoteinvestigator.com/2016/04/25/get/

1 Introduction

The current state of animal testing in preclinical safety, while long-established, faces increasing scrutiny due to ethical concerns and limitations in translating results to human-relevant outcomes (Hartung, 2013; Olson et al., 2000; Monticello et al., 2017). It is worth noting that *in vivo* studies have been used in all phases of pharmaceutical development programs for decades, without formal validation or qualification. The development and implementation of new approach methods (NAMs) such as *in vitro* or computational models in toxicology and safety assessment is of paramount importance for several compelling reasons. NAMs offer a greater ability to accurately predict the toxicity of chemicals and the efficacy of drugs in humans compared to animal testing methods as they closely mimic human physiological responses (Hartung, 2024a), thus promising better protection of human health. There is also a growing societal and scientific consensus on the need to reduce and, where possible, replace the use of animals in research (von Aulock et al., 2022). This shift addresses longstanding concerns about animal welfare and reflects a broader ethical imperative to minimize animal suffering in scientific pursuits. By providing alternatives to animal testing, NAMs align with these ethical principles while still advancing scientific knowledge and public health protection (Hartung, 2024b). Furthermore, NAMs have the potential to accelerate innovation in drug development, cosmetics, and industrial chemicals as these methodologies can often be implemented more rapidly and cost-effectively than traditional animal testing protocols (Meigs et al., 2018). This could also democratize research and development, enabling smaller companies and research institutions to engage in innovative work that might otherwise be prohibitively expensive.

Despite their potential, the implementation of NAMs faces several significant challenges. One of the primary hurdles is the complexity of validation. Validating these new methodologies can be a taxing process, particularly when it comes to simulating complex biological systems or predicting long-term health effects. The human body's intricacy and the multifaceted nature of many toxicological processes make it challenging to ensure that NAMs can reliably replicate or predict these phenomena. This complexity necessitates rigorous scientific scrutiny and extensive testing (Hartung et al., 2024b). Regulatory acceptance presents another significant challenge. Regulatory bodies must be confident that NAMs can provide results that are at minimum as reliable and predictive as traditional animal testing methods. This process can take years to complete. Lastly, the effective use of NAMs often depends on large datasets and collaborative efforts within the scientific community. Many NAMs rely on complex computational models or high-throughput screening methods that require substantial amounts of data to be effective. Gathering, standardizing, and sharing this data across different research institutions and companies can be challenging. It requires overcoming competitive barriers, establishing data sharing protocols, and ensuring data quality and compatibility. Furthermore, the interdisciplinary nature of many NAMs necessitates collaboration between experts in various fields, including biology, chemistry, computer science, and regulatory affairs. Overcoming these hurdles will require sustained effort, investment, and cooperation from various stakeholders in the scientific, regulatory, and industrial communities. Therefore, while the importance of developing and implementing NAMs is clear, addressing these challenges is crucial for their widespread adoption and success.

The validation of new approach methods (NAMs) for safety testing stands to benefit significantly from the concepts of translational medicine (TM)[2] and biomarkers, biomarker regulatory qualification, and the process and concept of TM in general. This integration is particularly relevant given the critical importance of preclinical safety in drug development, the safety testing of chemicals, and the growing need for alternatives to animal testing.

Biomarker identification and selection is crucial in TM but is not broadly applied in NAMs. However, NAMs can utilize biomarkers as predictive endpoints for safety testing or as surrogate indicators of the clinical efficacy of candidate drugs. Identifying and selecting relevant panels of biomarkers that are associated with specific toxicological profiles can provide a higher predictive power for NAMs. For example, consider a hepatocyte-based test system, for which a simplistic NAM so far might use viability as the endpoint. A set of complex transcriptomics endpoints could provide more information on the type of damage (e.g., cholestatic vs steatotic) and predict cellular perturbations with a higher sensitivity. Another example could be a test system based on neuronal networks, where early-generation NAMs measure structural defects (e.g., breaking neurites) but more sensitive and more discriminating information may be obtained using multi-electrode arrays that record changes in the electrical signaling within the network. Thus, biomarkers used as NAM endpoints can contribute to a deeper mechanistic understanding and serve as early indicators of potential adverse effects, guiding the development of targeted and relevant NAMs. NAMs that incorporate biomarkers with established translational relevance (Mattes and Walker, 2009; Mattes et al., 2010) can better predict human safety outcomes, instead of being optimized to predict animal test results (Sauer and Porter, 2018). This will be particularly important in the context of a Human Exposome Project (Hartung, 2023a; Sillé et al., 2024) to map the exposure side of diseases via mechanistic pathways.

The process of biomarker qualification, which involves establishing the biological and analytical validity of a biomarker for a particular context of use (CoU), can also be applied to NAMs. Qualification ensures that the biomarker used in a NAM is reli-

Abbreviations: AOP, adverse outcome pathway; BEST, Biomarkers, EndpointS, and other Tools; BQP, biomarker qualification program; CDER, Center for Drug Evaluation and Research; CoU, context of use; EMA, European Medicines Agency; FDA, Food and Drug Administration; IATA, integrated approaches to testing and assessment; iPSC, induced pluripotent stem cells; ITS, integrated testing strategy; KE, key event; MIE, molecular initiating events; NAMs, new approach methods; QIVIVE, quantitative *in vitro* to *in vivo* extrapolation; TM, translational medicine

[2] Not excluding by using the term translational medicine wider aspects of translational research and translational science: https://tri.uams.edu/news/translational-research-vs-translational-science-whats-the-difference/

Tab. 1: Key similarities and difference between translational medicine/biomarkers and the validation process for NAMs

Similarities:

- *Rigorous validation*: In both cases, a rigorous validation process is essential. Biomarkers need to be demonstrably linked to a specific disease state or treatment response, while alternative methods need to show they can reliably predict what would happen in traditional animal tests or ideally, in human biological pathways, e.g., via comparison to data from human clinical trials.
- *Reproducibility and specificity*: Both validated biomarkers and validated NAMs need to be reproducible (consistently produce the same results) and specific (accurately measure what they are intended to). This ensures they provide reliable information and can be used with confidence.
- *Addressing uncertainty*: Both approaches aim to reduce uncertainty in the transition from the controlled lab environment to real-world application. Biomarkers help make diagnoses or treatment decisions more precise, while validated NAMs offer a more accurate picture of how a substance might affect humans without animal testing.
- *Regulatory acceptance*: For both, gaining regulatory acceptance is a crucial step for widespread adoption. Regulatory bodies need to be convinced of the validity and reliability of these approaches before they can be routinely used in clinical practice or product development.
- *End goal*: TM/biomarkers aim to improve patient care by providing more personalized and precise diagnostic and treatment decisions. While NAM validation's original goal is ethical (to reduce or replace the use of animals), it similarly aims to improve standards of human safety during product development.

Differences:

- *Focus and scope*: TM/biomarkers are focused on identifying and using biological indicators that can aid in disease diagnosis, designing treatment strategies and predicting response, or monitoring disease progression. Biomarkers can come from a variety of sources, including genes, proteins, or even imaging. NAM validation is focused on developing and validating new approaches (e.g., complex *in vitro* models, computational simulations) to replace or complement the use of animals in research for testing safety and efficacy of drugs, chemicals, and other products.
- *Tools and techniques*: TM/biomarkers use a broad array of analytical tools to identify and measure biomarkers, including genomic and proteomic analysis, imaging techniques, and clinical assays. NAM validation currently employs a more specific set of tools, including cell-based assays (*in vitro* models), tissue engineering (like organs-on-chips), and advanced computer simulations. However, the diverse analytical tools harnessed in TM/biomarker can also be leveraged in future NAM development.
- *Path to adoption*: TM/biomarkers often involve the discovery and validation of novel biomarkers, along with their integration into drug development, clinical guidelines, and practice. NAM validation is primarily focused on demonstrating that the new method is equally or more reliable and relevant than existing animal models, and then gaining regulatory acceptance that allows them to be used in lieu of animal testing.

able, reproducible, and has a well-defined relationship with the toxicological endpoint of interest. Regulatory agencies are more likely to accept NAMs that utilize well-characterized and clinically relevant biomarkers. The measurement of qualified biomarkers in NAM studies supports incorporation of NAMs into regulatory decision-making processes. By leveraging the concepts of TM and biomarkers, their qualification, and the principles of TM, the validation of NAMs for safety testing can thus be facilitated.

TM and the validation process for NAMs share the core focus to better predict changes to human biological processes, one with the goal to assess medical interventions, the other with a focus on predicting the safety of chemicals. TM focuses on improving patient care, while alternative methods validation prioritizes replacing animal testing with ethically driven yet reliable methodologies. In essence, both tools aim to establish reliable and objective tools that can bridge the gap between research and practical applications that directly impact human health. Table 1 gives a breakdown of the similarities and differences.

This article aims to explore how TM and biomarkers in preclinical safety assessment can serve as supplements or alternatives to traditional animal testing, potentially revolutionizing the field. In turn, integrating these tools into NAMs can provide us with more efficient, ethical, and predictive safety testing paradigms. This includes how biomarker qualification offers opportunities to rethink and enrich the NAM validation process. While in the following we will focus mainly on the US Food and Drug Administration (FDA) Center for Drug Evaluation and Research (CDER) biomarker qualification process (BQP)[3] for clinical biomarkers, it is important to note that the European Medicines Agency (EMA) has similar processes (Bakker et al., 2022; Berman and Siegel, 2022).

2 The role of biomarkers in preclinical safety

Biomarkers have emerged as essential tools in the field of preclinical safety assessment[4] (for review see: Bleavins et al., 2010; Kumar and van Gool, 2013; FDA-NIH Biomarker Working Group, 2016; Horien, 2017; Califf, 2018; Gromova et al., 2020; Vlasakova et al., 2022; Bodaghi et al., 2023; Ahmad et al., 2023; Das et al., 2024), offering a means to detect, predict, and monitor potential effects of drugs. Biomarkers can provide valuable insights into the mechanisms of drug action or toxicity and help guide the development of safer and more effective therapies (Troth et al., 2019; Chen et al., 2018). To date, TM and biomarkers have focused mainly on the efficacy of drug action; the concept has been less commonly applied to assess drug adversity or to substances with other uses, e.g., industrial or food chemicals. Those biomarkers used to identify drug adversity (Sasseville et al., 2014) can be most easily translated to the field of safety testing, the primary realm of NAMs.

[3] https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/biomarker-qualification-program

[4] https://www.fda.gov/drugs/biomarker-qualification-program/biomarker-guidances-and-reference-materials

2.1 Types of biomarkers

Biomarkers can be classified in various ways. The BEST (Biomarkers, EndpointS, and other Tools)[4] resource goes into biomarker types quite extensively. According to their purpose, one can differentiate between biomarkers of disease progression, of differential diagnosis, and of treatment success. According to the technologies applied, biological, chemical, and physical approaches may be distinguished. Biological markers include molecular entities such as RNA, proteins, and metabolites and can be measured in various matrices such as blood, urine, or tissue, thus providing information about specific cellular processes or pathways affected by a drug or chemical. Chemical biomarkers are exogenous substances or their metabolites that can be used to assess exposure to a particular compound. These biomarkers can help determine the absorption, distribution, metabolism, and excretion of a drug or chemical in the body. Physical biomarkers include physiological parameters such as blood pressure, heart rate, or body temperature, which can be monitored to assess the overall health status of an organism and detect potential adverse effects.

2.2 The role of biomarkers in predicting drug safety and efficacy in preclinical animal tests

Several case studies demonstrate the successful application of biomarkers in preclinical safety assessment (Institute of Medicine, 2010; Schomaker et al., 2019). These biomarkers allow the prioritization of safer drug candidates for further development. For example, a panel of hepatic biomarkers, including alanine aminotransferase (ALT), aspartate aminotransferase (AST), and bilirubin, is routinely used to assess the hepatotoxic potential of compounds. Similarly, cardiovascular biomarkers such as troponin I and B-type natriuretic peptide (BNP) are used to evaluate the cardiotoxic effects of drug candidates (Sauer and Porter, 2018), and several kidney biomarkers, including kidney injury molecule-1 (KIM-1) and neutrophil gelatinase-associated lipocalin (NGAL), allow for early detection of kidney injury (Ozer et al., 2010; Bonventre et al., 2010; Chen et al., 2018; Troth et al., 2019; Sauer and Porter, 2018).

In exploratory or investigative toxicology (Beilmann et al., 2019), biomarkers play an important translational role by connecting biomarker changes with histopathological injury and, therefore, provide a more specific and sensitive tool to detect and monitor drug-induced organ injury, often before overt clinical signs or symptoms appear. By measuring specific biomarkers that are mechanistically linked to the therapeutic target or pathway, the pharmacodynamic effects of a drug and its potential efficacy can be assessed. The integration of safety and efficacy biomarkers in preclinical studies thus provides a comprehensive understanding of a drug's risk-benefit profile, enabling informed decision-making in the drug development process.

2.3 The relevance of biomarkers in detecting and predicting safety issues *in vitro*

A conceptual framework for incorporating biomarkers in *in vitro* systems for toxicological risk assessment as alternatives to animal testing is discussed in Blaauboer et al. (2012). The authors define biomarkers of toxicity as providing quantitative information that is mechanistically relevant to and predictive of an adverse effect *in vivo*. They emphasize the importance of considering biokinetics both *in vitro* (e.g., free vs nominal concentrations) and for quantitative *in vitro* to *in vivo* extrapolation (QIVIVE) and discuss the challenge of distinguishing between adaptive and adverse effects when analyzing *in vitro* toxicity data. The proposed framework includes steps like evaluating exposure scenarios, structural properties, pharmacokinetic behavior, and using *in vitro* effect batteries. For data interpretation, the paper suggests modeling concentration-effect relationships and determining appropriate points-of-departure for risk assessment. It highlights the need for QIVIVE using physiologically based biokinetic models. The authors also discuss challenges in implementing this approach, including lack of standardization and limited experience with non-animal-based risk assessment. They emphasize the need for more research to build experience with chemical risk assessment using *in vitro* biomarker data and the proposed framework. These key points align well with and support many of the concepts discussed here on leveraging biomarkers and TM for NAMs. In the decade since the publication of this framework, many publications have expanded and built upon the proposed concepts (Krewski et al., 2014; Burgdorf et al., 2019; Parish et al., 2020; Magurany et al., 2023).

A fundamental question is the difference between an ordinary *in vitro* endpoint and an *in vitro* biomarker. Based on the BEST[4] resource and general understanding in the field, there are some key differences between an *in vitro* endpoint in a NAM and a biomarker:

a) *Definition:* A biomarker is defined as "*a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or biological responses to an exposure or intervention, including therapeutic interventions*." A NAM endpoint is not explicitly defined in the BEST[4] resource, but generally refers to a measurement or outcome observed in a laboratory setting outside of a living organism.
b) *Scope:* Biomarkers encompass a wide range of measurements, including molecular, histologic, radiographic, or physiologic characteristics. They predict human outcomes. NAM endpoints are typically more specific to laboratory-based assays and may include measures like enzyme activity, cell viability or protein expression in cultured cells. They traditionally predict outcomes in an animal experiment, although increased emphasis is placed on human biological relevance[5].
c) *Application:* Biomarkers are often used to inform decisions about clinical care, drug development, or health risk assessment in living subjects. NAM endpoints are primarily used in early-stage research, drug screening, or mechanistic studies that do not directly involve living subjects.
d) *Regulatory context:* Biomarkers have specific regulatory definitions and can be used in various ways in drug development and approval processes, including as surrogate endpoints. NAM endpoints generally do not have the same regulatory sta-

[5] https://ntp.niehs.nih.gov/sites/default/files/2024-03/VWG_Report_27Feb2024_FD_508.pdf

tus as biomarkers and are typically used earlier in the research and development process, although there are examples, such as skin sensitization, where NAMs have demonstrated superior performance to animal studies and can be used in regulatory decision making, e.g., OECD TG 497 (OECD, 2023).

e) *Validation requirements:* Biomarkers often require extensive validation, including analytical and clinical validation, especially if they are to be used in regulatory decision-making. NAM endpoints may require method validation, but the process is typically less rigorous than for biomarkers intended for clinical use. Where NAMs are to be considered as drug development tools for regulatory decision-making, validation will require the respective confidence.

Biomarkers have been successfully applied across various omics disciplines in preclinical safety assessment. In toxicogenomics, genomic biomarkers have been used to identify gene expression signatures associated with specific toxicological outcomes, such as genotoxicity or carcinogenicity. These gene expression profiles can be used to screen compounds for their toxic potential, e.g., to cause DNA damage or induce tumor formation. In toxicoproteomics, protein biomarkers have been employed to assess the effects of drugs or chemicals on cellular signaling pathways and identify potential mechanisms of toxicity. For instance, changes in the expression of heat shock proteins or oxidative stress markers can indicate cellular stress responses and guide the development of safer compounds. Metabolic biomarkers, studied in toxicometabolomics (Bouhifd et al., 2013; Ramirez et al., 2013; Sillé and Hartung, 2024), have been used to evaluate the impact of drugs or chemicals on cellular metabolism and identify potential metabolic disruptions. By quantifying changes in (patterns of) specific metabolites, researchers can gain insights into the metabolic pathways affected by a compound and correlate its potential toxicity.

It is worth noting that some *in vitro* endpoints, especially in human-relevant test systems such as microphysiological systems (MPS), could potentially become biomarkers if they are found to reliably indicate a biological process or response in human organisms and undergo appropriate validation. However, not all *in vitro* endpoints will become biomarkers, and not all biomarkers are derived from *in vitro* measurements.

By measuring specific biomarkers *in vitro*, e.g., in multi-organ-on-a-chip systems, potential organ-specific toxicities such as liver, kidney, or cardiac damage can be observed. Further, the dose-response relationships of a drug or chemical can be identified for each tissue. Multi-organ chips also enable studies of off-target effects and of toxicities induced by metabolic byproducts produced in one organ on other organs (e.g., Chang et al., 2017).

Biomarkers also allow for the prediction of long-term safety outcomes, as changes in certain biomarkers may be indicative of chronic toxicity or delayed adverse effects (Schomaker et al., 2019).

2.4 Limitations and challenges of biomarker discovery and validation

Despite the numerous advantages of biomarkers in preclinical safety assessment, there are several limitations and challenges associated with their discovery and validation (Davis et al., 2020). Some biomarkers may lack specificity, as they can be associated with multiple biological processes or pathways, complicating the interpretation of their changes. Analytical challenges in biomarker measurement, such as assay sensitivity, reproducibility, and standardization, also pose significant hurdles in biomarker validation. Regulatory acceptance therefore depends on qualification of biomarkers, which requires rigorous validation and documentation to ensure their reliability and relevance.

2.5 Future perspectives and opportunities for preclinical biomarkers

The field of biomarker-based preclinical safety assessment is rapidly evolving. High-throughput technologies, such as next-generation sequencing and mass spectrometry, have revolutionized the discovery of novel biomarkers, enabling the identification of more sensitive and specific indicators of toxicity. The integration of genomics, proteomics, metabolomics, and other omics disciplines provides a systems biology level understanding of the biological responses to drugs or chemicals, facilitating the development of comprehensive biomarker panels. Moreover, the development of novel biomarker-based assays and platforms, such as organ-on-a-chip systems and 3D cell culture models, offers new opportunities for linking preclinical biomarkers with safety/efficacy and the assessment of systemic or organ-specific toxicities. Collaborative efforts among academia, industry, and regulatory agencies are crucial for the validation and standardization of biomarkers, ensuring their reproducibility and reliability in preclinical safety assessment.

3 Translational medicine – Bridging the gap between laboratory and clinic

TM is a multidisciplinary field that aims to bridge the gap between preclinical research and clinical practice, facilitating the translation of scientific discoveries into effective and safe therapies for patients (Wehling, 2015). TM also emphasizes the importance of incorporating clinical insights and patient perspectives into preclinical research. By engaging with clinicians and patients, researchers can gain a better understanding of the unmet medical needs, disease burden, and treatment challenges associated with a particular condition (Sung et al., 2003; Brown et al., 2020; Kimmel et al., 2019). This knowledge can inform the design of preclinical studies.

In the context of drug development, TM plays a crucial role in ensuring that preclinical findings are relevant and applicable to human health. Preclinical studies conducted in animal models provide valuable insights into the safety and efficacy of a drug candidate. However, the translation of these findings to clinical settings can be challenging due to differences in species, disease pathology, and pharmacokinetics (Mankoff et al., 2004; Tsaioun et al., 2016; Davis et al., 2020).

One key aspect of TM is the use of human-derived cells, tissues, and organs in preclinical research. By using human-relevant models, researchers can better understand disease mechanisms and treatment effects in a human-specific context. For example, the

use of human induced pluripotent stem cells (iPSCs) has revolutionized drug discovery and toxicity testing (Roth et al., 2021), allowing for the generation of patient-specific cell lines that capture the genetic and phenotypic diversity of human populations (for the example of pain therapeutics: Davis et al., 2020). Additionally, the use of patient-derived samples, such as biopsies or blood samples, can provide valuable insights into disease mechanisms and treatment responses, guiding the development of more effective and targeted therapies (El-Achkar et al., 2024). As TM continues to evolve, the integration of advanced technologies such as omics, imaging, and computational modeling will further enhance our ability to understand disease mechanisms, predict treatment responses, and develop targeted therapies that benefit patients.

Biomarkers are indispensable in the field of TM. Biomarkers are measurable indicators of biological processes, disease states, or treatment responses that can be used to assess the safety and efficacy of a drug candidate (Bleavins et al., 2010). For example, the use of imaging biomarkers, such as positron emission tomography (PET) and magnetic resonance imaging (MRI), has enabled the non-invasive assessment of drug distribution, target engagement, and treatment response in both preclinical and clinical studies (for the example of cancer: Luciano et al., 2011). The integration of biomarkers into clinical practice is a key factor in the realization of personalized medicine, which promises to revolutionize healthcare by providing tailored and effective treatments for patients (Hood et al., 2004; Hendrikse et al., 2022; Song and Dobbin, 2022; Wang and Ward, 2012; Bravo-Merodio et al., 2019; Wagner and Srivastava, 2012).

One major challenge is the biological variability and the dynamic range of biomarkers across populations. Biomarker responses can be influenced by factors such as age, sex, genetic background, and environmental conditions, among others, making it challenging to establish robust and reproducible biomarker thresholds and signatures. Translational approaches aim to address these challenges by incorporating human-relevant models, biomarkers, and other endpoints into preclinical studies[6]. Increasingly, adverse outcome pathways (AOP) (Leist et al., 2017) are utilized in the drug development context for safety (Hartung, 2017) or for disease characterization (Clerbaux et al., 2022).

Beside the biomarkers of toxicity, which are most relevant for NAMs, there are several other types of clinical biomarkers (Box 1), discussed here only in general terms.

Box 1: Types of clinical biomarkers

Translational medicine leverages a variety of biomarker types, including genomic, proteomic, metabolomic, and imaging biomarkers. The integration of these diverse modalities through systems biology approaches enhances the predictive power and clinical utility of biomarkers. By combining data from multiple sources, a more comprehensive and accurate picture of a patient's condition can be constructed, leading to better-informed clinical decisions and improved patient outcomes (Wagner et al., 2012).

a) Disease diagnosis and prognosis

Biomarkers are instrumental in the early detection and diagnosis of diseases. They offer a measurable reflection of biological states and associated perturbations, which can be indicative of underlying pathologies even before clinical symptoms manifest. This early detection is crucial, as it allows for the initiation of treatment at a stage when it is likely to be more effective, potentially altering the course of the disease (Hendrikse et al., 2022). Furthermore, biomarkers can prognosticate the likely trajectory of a disease (Murad and Melamud, 2022), providing clinicians with insights into the expected progression and enabling healthcare providers to tailor treatment strategies to the individual patient, optimizing outcomes and improving quality of care (Hendrikse et al., 2022).

b) Treatment selection and monitoring

In the era of precision medicine, biomarkers serve as the cornerstone for selecting targeted therapies that are tailored to the molecular profile of an individual's disease and potentially their personal genetic makeup (Serelli-Lee et al., 2022). This approach ensures that patients receive treatments that are most likely to be effective for their specific condition, thereby maximizing therapeutic efficacy and minimizing the risk of adverse effects (Song and Dobbin, 2022). Additionally, biomarkers facilitate the real-time monitoring of treatment responses and disease progression. This dynamic monitoring allows for adjustments in therapy to be made in a timely manner, ensuring that patients receive the most appropriate and effective care throughout their treatment course (Song and Dobbin, 2022).

c) Drug development and clinical trials

The drug development process benefits from the use of biomarkers. They aid in the identification of therapeutic targets and enable the screening of drug candidates in preclinical models. Biomarkers also serve as surrogate endpoints in clinical trials, providing a means to assess treatment response without the need for biopsies or long-term clinical outcome studies. This surrogate role can expedite the drug development timeline and reduce associated costs, making the process more efficient and potentially bringing effective treatments to patients sooner (Wang and Ward, 2012).

d) Mechanistic insights and disease understanding

By identifying the molecular pathways involved in the initiation and progression of a disease (Robinson et al., 2013; Owen et al., 2023), as well as the response to treatment, biomarkers contribute to a deeper understanding of disease pathophysiology. This knowledge is invaluable for the development of novel therapeutic strategies and the identification of new drug targets, ultimately leading to more effective treatments and interventions (Bravo-Merodio et al., 2019).

[6] https://aspe.hhs.gov/sites/default/files/private/pdf/260031/FinalBiomarkersReport.pdf

4 Current advances in non-animal preclinical models vs challenges in the validation process of new approach methods for safety testing

Significant challenges hinder the widespread validation and adoption of NAMs. Foremost among these is the inherent complexity of replicating the intricate biological systems of an organism within simplified NAMs. Moreover, the validation process is hampered by the limited availability of reliable and comprehensive human reference data, as animal test results can be inconsistent or irrelevant to human biology (Olson et al., 2000; Monticello et al., 2017; Ewart et al., 2022). Therefore, the comparison of results from human cell-based NAMs with results from animal tests may not fully align although the NAM may better predict the human response.

Beyond the scientific challenges, regulatory acceptance of NAMs poses another barrier. Regulatory agencies often have long-established protocols and standards built on animal data that may not be fully transferable or appropriate for novel NAMs. Accepting NAMs requires rigorous validation, a process that is arduous, expensive, and time-consuming. Smaller companies or research laboratories may lack the financial means or specialized expertise to drive the development and adoption of these advanced techniques. Continued investment in alternative methodologies is crucial to ensure they can demonstrably and reliably replace traditional animal testing paradigms. In 2024, the US FDA established a pilot program for regulatory acceptance of NAMs that address regulatory decision-making questions[7]. The first method has been accepted into this program[8].

Box 2: Key concepts and trends in the validation of NAMs

International cooperation and harmonization
International cooperation, such as the International Cooperation on Alternative Test Methods (ICATM)[9], plays a crucial role in the validation and regulatory acceptance of NAMs. This facilitates the exchange of information, best practices, and harmonization of validation processes across different countries and regulatory bodies (Stucki et al., 2022; Mondou et al., 2021). Efforts are made to align with the Organisation for Economic Co-operation and Development (OECD) guidelines and test methods, which support the mutual acceptance of data among member countries (Stucki et al., 2022).

Scientific confidence frameworks
Developing scientific confidence frameworks for NAMs is essential for their acceptance and integration into regulatory decision-making[8,10,11]. These frameworks help in evaluating the reliability, relevance, and adequacy of NAMs for specific regulatory purposes (van der Zalm et al., 2022). The frameworks often include criteria for internal and external validity, experimental variability, and technical characterization, and the use of structured approaches for evaluating the performance of test methods[12,13].

Fit-for-purpose validation
The concept of "fit-for-purpose" validation (Cummings et al., 2010; Lee et al., 2006) is gaining traction, emphasizing that the validation process should ensure NAMs are suitable for their intended use. This approach recognizes that NAMs may provide different but equally valuable information compared to traditional animal tests and may offer more human-relevant insights (van der Zalm et al., 2022). The validation process should be flexible, considering the specific context and purpose of the NAM[14].

Integrated testing strategies and integrated approaches to testing and assessment
ITS/IATAs represent a holistic approach to chemical safety assessment, combining data from various NAMs, including *in vitro*, *in chemico* and *in silico* methods. IATAs aim to provide a comprehensive understanding of a substance's hazard potential through a weight-of-evidence approach (Stucki et al., 2022; Caloni et al., 2022).

Regulatory acceptance and implementation
For NAMs to be effectively integrated into regulatory frameworks, clear guidance on their use and acceptance is crucial. Regulatory agencies are working towards providing such guidance and establishing processes for the inclusion of NAMs in risk assessments and decision-making[15]. This includes efforts to update regulatory requirements to accommodate NAM data and to develop criteria for their acceptance[16].

7 https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/innovative-science-and-technology-approaches-new-drugs-istand-pilot-program
8 https://force-dsc.my.site.com/ddt/s/ddt-project?ddtprojectid=173
9 https://bit.ly/3XYIP4T
10 https://echa.europa.eu/documents/10162/21838212/scientific_ws_proceedings_en.pdf/a2087434-0407-4705-9057-95d9c2c2cc57
11 https://www.fda.gov/food/toxicology-research/new-approach-methods-nams
12 https://nap.nationalacademies.org/read/26906/chapter/7
13 https://nap.nationalacademies.org/read/26496/chapter/1
14 https://www.rivm.nl/sites/default/files/2022-10/22402706_013889_br%20vormgeving%20landschap%20eng_v3_tg.pdf
15 https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/alternative-test-methods-and-strategies-reduce
16 https://ntp.niehs.nih.gov/whatwestudy/niceatm/resources-for-test-method-developers/method-developers-forums

Challenges and future directions
Despite significant progress (Mahony, 2019), challenges remain in the wider application and acceptance of NAMs[17], including validation hurdles, the need for more comprehensive databases of NAMs, and the integration of NAM data into regulatory frameworks. Future directions involve addressing these challenges, further developing and refining NAMs, and enhancing international collaboration to facilitate their global acceptance (Stucki et al., 2022)

These efforts aim to overcome existing challenges and pave the way for the broader implementation of NAMs in chemical safety assessments[12,18,19,20] (Stucki et al., 2022; Mondou et al., 2021; van der Zalm et al., 2022; Schmeisser et al., 2023). Continued investment in research[21], development of standardized validation frameworks, and open communication among stakeholders will be essential for advancing the validation and acceptance of NAMs for safety testing.

The validation of NAMs[5], including MPS (Marx et al., 2016, 2020; Edington et al., 2018), artificial intelligence (AI) in toxicology (Kleinstreuer and Hartung, 2024), and integrated testing strategies (ITS) (Caloni et al., 2022), thus presents a complex set of challenges stemming from their innovative nature, their technical complexity, the attempt to supersede traditional "gold standard" animal models, and the regulatory environment in which they are to be applied.

4.1 Microphysiological systems

Microphysiological systems (MPS) are microfluidic devices capable of emulating human (or any other animal species') biology *in vitro* at the smallest biologically acceptable scale, defined by purpose. The application of fluid flow (dynamic) for the physiological nutrition of the tissues and the creation of microenvironmental biomolecular gradients and relevant mechanical cues (e.g., shear stress) is a major aspect of these systems, differentiating them from conventional (static) cell and tissue cultures (Marx et al., 2020).

a) *Reproducibility and standardization:* A significant challenge for MPS is ensuring reproducibility and standardization across different laboratories and platforms (Malik et al., 2021). MPS are intricate systems that can vary widely in design, materials, and operating conditions, which can affect the results obtained. Establishing standardized protocols and benchmarks is essential for the validation and broader acceptance of MPS (Ewart et al., 2017; Hargrove-Grimes et al., 2021, 2022; Kopec et al., 2021; Mansouri et al., 2024; Pamies et al., 2022, 2024).
b) *Physiological relevance:* Another challenge is ensuring that MPS accurately replicate human physiology and disease states. This includes the integration of multiple cell types, the recapitulation of the 3D architecture of tissues, and the simulation of dynamic physiological processes. Demonstrating the physiological relevance of MPS models requires comprehensive characterization and comparison with *in vivo* data (Hargrove-Grimes et al., 2021; Mansouri et al., 2024). Their combination with omics approaches is promising (Del Giudice et al., 2023; Schwartz et al., 2015).
c) *Regulatory acceptance:* For MPS to be used in regulatory decision-making, they must be accepted by regulatory agencies. This requires the development of guidelines for the use of MPS in specific regulatory contexts, including criteria for validation and the types of data that MPS need to generate. Engaging with regulatory agencies early in the development process is crucial for understanding regulatory requirements and facilitating the acceptance of MPS (Tagle, 2019; Kopec et al., 2021; Mansouri et al., 2024).

For example, the IQ MPS Affiliate[22] has had significant interactions with the US FDA focused on MPS and spheroids[23]. There is an enormous potential to apply validation concepts here for a given CoU in a pharma company.

4.2 Artificial intelligence in toxicology

We have recently discussed the advances of AI in this series (Hartung, 2023b) and elsewhere (Hartung, 2023c; Kleinstreuer and Hartung, 2024). Here only some major challenges shall be reiterated:

a) *Data quality and availability:* AI models are only as good as the data they are trained on. A major challenge is the availability of high-quality, annotated datasets for training and testing AI models. In toxicology, this includes data from *in vitro* assays, animal studies, and human exposure data. Ensuring the quality, completeness, and representativeness of these datasets is critical for the development of reliable AI models (Kleinstreuer and Hartung, 2024).
b) *Interpretability and transparency:* AI models, especially deep learning models, are often criticized for being "black boxes" that provide little insight into how they arrive at their predictions. For regulatory purposes, it is important that AI models are interpretable and transparent, so that their predictions can be understood and trusted (Lin and Chou, 2022; Kleinstreuer and Hartung, 2024; Tonoyan and Siraki, 2024)
c) *Validation and benchmarking:* Validating AI models in toxicology involves demonstrating their predictive performance and reliability across a range of chemicals and biological end-

17 https://www.nationalacademies.org/our-work/variability-and-relevance-of-current-laboratory-mammalian-toxicity-tests-and-expectations-for-new-approach-methods--nams--for-use-in-human-health-risk-assessment
18 https://ntp.niehs.nih.gov/whatwestudy/niceatm/natl-strategy
19 https://www.canada.ca/en/health-canada/services/chemical-substances/fact-sheets/use-new-approach-methods-risk-assessment.html
20 https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/strategic-vision-adopting-new-approach-2
21 https://commonfund.nih.gov/complementarie
22 https://iqconsortium.org
23 https://iqconsortium.org/news/iq-mps-affiliate

points. This requires the development of benchmark datasets and performance metrics. Comparing AI models to traditional toxicological methods and demonstrating their added value is also important for their acceptance (Kleinstreuer and Hartung, 2024).

4.3 Integrated testing strategies

ITS, aka integrated approaches to testing and assessment (IATA), involve the integration of data from various sources, including *in vitro* assays, *in silico* models, and chemical structure-activity relationships.

a) *Integration of diverse data types:* A challenge is developing methodologies for integrating these diverse data types in a coherent and scientifically sound manner (Hartung et al., 2013a; Rovida et al., 2015; Nendza and Ahlers, 2022).
b) *Regulatory frameworks:* For ITS to be used in regulatory decision-making, they must fit within existing regulatory frameworks[24]. This involves developing guidelines for the design, validation, and use of ITS in chemical risk assessment. It also requires demonstrating that ITS can provide equivalent or better protection for human health and the environment compared to traditional testing methods (Hartung et al., 2013a).
c) *Uncertainty analysis:* ITS inherently involve a degree of uncertainty due to the use of data from different sources and models with varying degrees of complexity. Developing methods for quantifying and communicating the uncertainty associated with ITS predictions is important for their acceptance and use in decision-making (Caloni et al., 2022).

In conclusion, the validation of MPS, AI, and ITS in toxicology presents a range of technical, scientific, and regulatory challenges. Addressing these challenges requires collaborative efforts from all users (researchers, method developers, industry scientists, and regulatory agencies) (Ball et al., 2022). Through such collaboration, it is possible to advance the validation and acceptance of these innovative methodologies, ultimately enhancing the efficiency and effectiveness of chemical safety assessment (Hartung and Leist, 2008; Leist et al., 2008; Tran et al., 2023).

4.4 Use of biomarkers to improve the applicability of adverse outcome pathways in hazard predictions

In the context of AOPs (Leist et al., 2017), biomarkers can take several roles. The different situations and the respective definitions of biomarkers are illustrated here. The AOP concept has an important role for NAM-based safety assessment, as the theoretical construct of an AOP can be used to translate molecular changes triggered by a toxicant to adverse outcomes (AO, i.e., apical endpoints) observed in humans. The pivotal role of this concept is linked to the facts that (i) apical endpoints can usually not be observed/assessed in NAM, and that (ii) NAM usually measure mechanistic changes historically not used for hazard characterization and classification in traditional toxicology. In a very wide sense, one may envisage AOPs as tools to relate biomarkers to AO (Fig. 1).

Thus, the first of the three main AOP perspectives on biomarkers is that many NAM endpoints may be considered biomarkers, and that an AOP relates them, in a predictive way, to an AO that is observable in an organism. Many NAMs measure molecular initiating events (MIE) or key events (KE) of AOPs. In the context of NAM optimization or validation, the biomarker endpoint may be chosen to be particularly human-relevant. For instance, cell viability is an endpoint that can be assessed by many analytical methods. However, transaminase release from hepatocyte cultures or altered cellular chromatin structure are endpoints that are closer to observations made in animal models or humans than methods used to measure viability, e.g., changed impedance values or an altered reduction of a redox dye (e.g., resazurin).

The second and third AOP perspectives on biomarkers are different and probably more important. They refer to indirect measurements of MIE and KE. In these situations, the biomarker is used as an indirect approach to quantify MIE/KE activation. This is analogous to the clinical use of biomarkers, i.e., to indirectly measure a pathological event that is not easily accessible for direct measurements.

In the case of MIE, there are often situations, where the AOP starting from a hypothetical MIE is known, but the assay (NAM) to assess the interaction of a chemical with the MIE is expensive, complicated or unreliable. In such cases, an indirect assessment, i.e., the measurement of a biomarker closely correlated to the MIE, may be beneficial. A typical example is AOP:3[25], which starts with binding of a toxicant to complex I (c-I) of the mitochondrial electron transport chain. Such assays are not easily available. Instead, one may use a stop of the electron flow through the electron transport chain as biomarker. This can be easily assessed. If one assumes that compounds that bind to c-I will also block its function, then the biomarker correlates well with the MIE. Another example may be the block of aromatase as MIE. This is difficult to measure directly. A suitable biomarker would be an altered production of some steroid hormones. Similarly, activation of the A_h receptor is difficult to quantify directly. Instead, one may measure a closely related event, e.g., induction of Cyp1A enzymes, as biomarker.

Concerning KE, it may not be immediately evident why a biomarker measurement is required. KE are per definition measurable (directly). However, there are indeed several situations, where an indirect assessment via a biomarker may be useful. This is linked to the fact that KE can be relatively large "bins" of biological processes. For instance, the KE "mitochondrial dysfunction" or "disturbed synapse formation" include dozens or even hundreds of biological processes and functions. (i) Sometimes a complex biomarker may assess such a KE more comprehensively than a KE-directed NAM that measures only one of many aspects. For instance, some toxicogenomic modules or "footprints" that integrate different aspects of a large KE may be useful. Another example is oxidative stress, which may require the assessment of many different reactive oxygen species and the cell antioxidant status by multiple NAM. Instead, Nrf-2 activation may be a biomarker that

[24] https://echa.europa.eu/documents/10162/17224/information_requirements_r7b_en.pdf/1a551efc-bd6a-4d1f-b719-16e0d3a01919
[25] https://aopwiki.org/aops/3

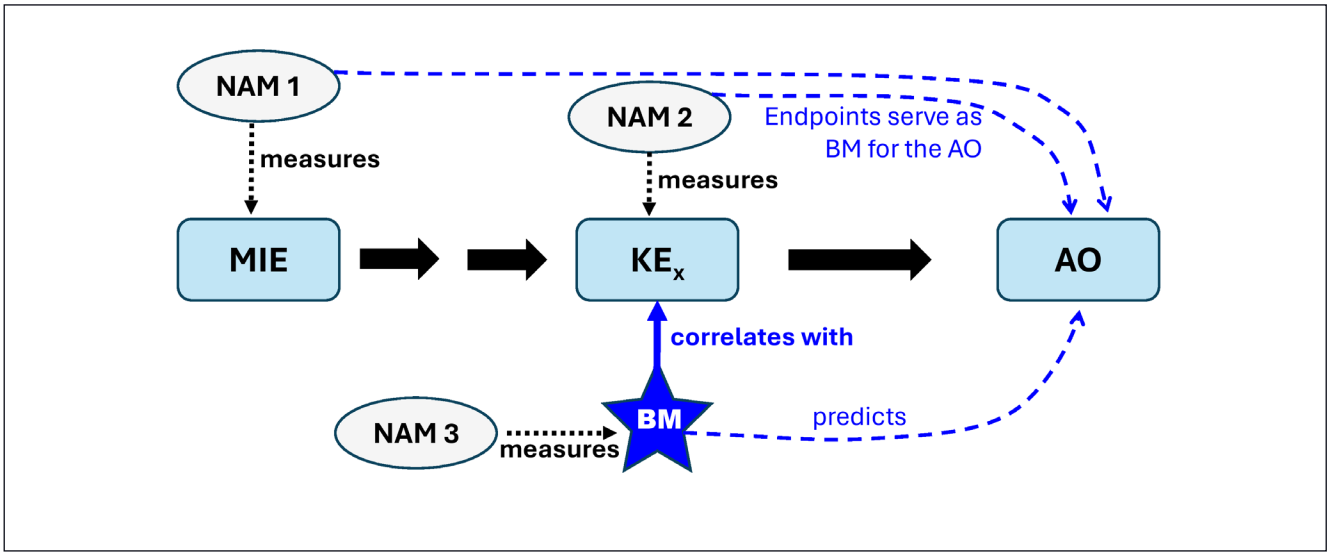

Fig. 1: Biomarkers in the context of AOP
An exemplary AOP with two of its key events (KE) and the adverse outcome (AO) is shown. Two different roles of biomarkers are exemplified. NAM 1 measures interference of test chemicals with the molecular initiating event (MIE). If the AOP is well-constructed and valid, then the results (endpoints) of NAM 1 serve as biomarker for the capacity of the test items to trigger the AO. NAM 2 measures a key event that is further downstream (KEx). The endpoint of NAM 2 may also serve as biomarker for the AO. As KEx is more upstream than the MIE, it may be better correlated to the AO than the MIE. The results of NAM 2 would therefore be a better biomarker than those of NAM 1, but this is not always the case. Some features of KEx may be difficult to assess. In such a case, activation of KEx may be assessed indirectly, i.e., via a biomarker that correlates with the triggering of KEx. NAM 3 may be used to measure this biomarker. As the biomarker correlates with the KE, it can be assumed to also predict the AO.

integrates most of these processes, though it is only an indirect proxy of the KE. (ii) Sometimes, a reverse situation may occur. Out of a "large" KE, only one aspect may translate to the AO. For instance, "synaptic dysfunction" has many consequences, some leading to certain AO, some not. One biomarker may be altered long-term potentiation, and this may correlate well with cognitive dysfunction as AO. Other test endpoints, e.g., altered catecholamine neurotransmitter loading into vesicles, may not correlate to this AO.

4.5 Integration of biomarkers to enhance the predictive power of NAMs

The integration of biomarkers into NAMs has significantly enhanced their predictive power:

a) In vitro *systems:* Biomarkers can be used to monitor cellular responses to toxicants in real-time. For example, measuring the release of specific enzymes or changes in gene expression can provide early indicators of cellular damage or stress before cytotoxicity as the most common endpoint is reached. Changes in mitochondrial membrane potential or ATP levels can be early indicators of mitochondrial dysfunction and cellular stress. Fluorescent dyes or luminescence-based assays can be used to monitor these parameters in real-time; activation of stress-responsive transcription factors like Nrf2 or heat shock factors can be monitored using reporter gene assays as early indicators of cellular stress responses. These biomarker approaches allow for real-time, sensitive detection of pre-cytotoxic cellular responses to toxicants, providing valuable information on mechanisms of toxicity and earlier endpoints for safety assessment of chemicals and drugs.
b) *Organ(s)-on-chips:* Biomarkers can be used to assess organ-specific toxicity. For instance, measuring troponin levels in a heart-on-a-chip model can indicate cardiotoxicity, while measuring albumin production in a liver-on-a-chip can indicate hepatotoxicity. This is particularly informative in multi-organ-on-chip systems.
c) *Computational models:* Biomarker data can be incorporated into computational models to improve their predictive accuracy. For example, toxicogenomic biomarkers can be used to develop more sophisticated quantitative structure-activity relationship models. Computational models can also be employed for biomarker discovery (Skolariki et al., 2023).

Some non-comprehensive examples of the application of biomarkers in drug and chemical safety assessments:

a) *Liver toxicity assessment:* A study by Ewart et al. (2022) used a liver-on-a-chip model integrated with biomarkers of liver function (e.g., albumin production, bile acid secretion) to assess drug-induced liver injury. The model successfully predicted the hepatotoxicity of several drugs that had caused liver damage in human patients despite having passed animal testing.
b) *Cardiotoxicity screening:* In the field of cardiotoxicity testing, human cell-derived cardiomyocyte iPSCs have been used for over a decade and have widespread FDA consideration (Guo et al., 2013; Yang et al., 2022; Pang et al., 2024). One nota-

ble example used cardiac troponin I release as a biomarker to demonstrate the response of a heart-on-a-chip model to doxorubicin, a known cardiotoxic drug (Takeda et al., 2018). The CIPA initiative[26] aims to engineer an assay for assessment of the proarrhythmic potential of new drugs.

c) *Nephrotoxicity prediction:* Kidney-on-a-chip models have been shown to integrate various biomarkers of kidney function and injury to assess drug-induced nephrotoxicity (Wilmer et al., 2016; Nguyen et al., 2023).
d) *Neurotoxicity assessment:* Schwartz et al. (2015) used a brain-on-a-chip model combined with electrophysiological biomarkers to assess neurotoxicity. The model was able to detect subtle changes in neuronal function induced by various neurotoxicants, demonstrating its potential for early detection of neurotoxic effects.
e) *Pesticide risk characterization:* Marciano et al. (2024) measured levels of triazole fungicides and associated biomarkers of health effects in occupationally and environmentally exposed human populations. Using *in vitro* high throughput systems and QIVIVE, they demonstrated that the molecular perturbations in NAMs, including liver enzyme expression, steroid hormone metabolism, and cellular stress properties, directly correlated with the human biomarker at equivalent exposures, highlighting human health risk and actionable mitigation measures.

These examples demonstrate the potential of NAMs integrated with biomarkers to provide valuable insights into drug safety and chemical toxicity. However, it is important to note that while these models show promise, they are still evolving. Challenges remain in terms of standardization, validation, and regulatory acceptance. The next chapter will delve into these challenges and discuss ongoing efforts to address them.

5 What the validation process for alternative methods can learn from translational medicine and biomarkers

NAMs face hurdles related to biological complexity, limited human reference data, species differences, regulatory acceptance, and developmental costs. Despite these challenges, the ethical imperative, potential for greater accuracy in predicting human safety, and advancements in technology drive the continued pursuit of reliable alternatives. Interestingly, the field of TM, which focuses on identifying and utilizing biomarkers to improve patient outcomes, offers a compelling parallel. The well-established qualification processes for biomarkers offer a valuable framework that can inform and potentially accelerate progress in validating NAMs:

a) *Adapting frameworks:* TM's robust frameworks for biomarker validation can be adapted to ensure consistency, reproducibility, and streamline the path to regulatory acceptance for NAMs.
b) *Context of use:* The critical concept in biomarker validation is how the biomarker will be utilized and what the question of interest is. The question of interest drives the CoU statement and is closely tied to "fit for purpose" applications. This is the bedrock lesson from biomarker regulatory qualification (US FDA, EMA or PMDA) – a tool is qualified or accepted for a specific use. The use drives the complexity of validation (analytical and clinical). Of course, NAM developers, vendors, and champions may rather want universal validation/qualification, but the record in pharma internal use and regulatory qualification would indicate that the CoU determines the validation needed for qualification/regulatory acceptance.
c) *Prioritizing clinical relevance:* Like a successful biomarker qualification, validation should demonstrate how a NAM can predict human responses (safety or efficacy) rather than mimic animal test results. Notably, to date, only prognostic and safety biomarkers have been qualified; there are no diagnostic/surrogate endpoint qualifications yet. Similarly, validation of NAMs should demonstrate how the new methods will improve the safety and efficacy testing of drugs or products for humans. The use of translational biomarkers as a potential approach for "validating" safety NAMs encourages the NAM community to embrace endpoints that are clinically relevant. And while the use of NAMs for a regulatory-approved clinical safety assessment is a goal, their use in regulatory approval for clinical efficacy is already here (for example: Ratner, 2017; Costa et al., 2022; Weaver et al., 2022)! Note also the decades-long use of NAMs in assessing drug metabolism; these studies are now essential elements of an investigational new drug (IND) package (for example: Hewitt et al., 2001).
d) *Collaboration as a catalyst:* Increased collaboration among developers of alternative methods, toxicologists, and regulatory bodies can accelerate progress and ensure methods address all necessary concerns. Sharing the risk and cost associated with developing new methods creates an environment that will expedite advancements and improve implementation. In the TM/biomarker field, public-private partnerships such as the Biomarkers Consortium[27] have dramatically impacted progress.
e) *Demonstrating impact:* The success of a biomarker is often measured by its real-world impact on improving patient care and outcomes (Troth et al., 2019; Chen et al., 2018). The validation process needs to clearly show how NAMs will enhance the safety and efficacy testing of products for humans.

By incorporating these lessons from TM, the validation process for NAMs can become more efficient and rigorous, and ultimately lead to the development of reliable replacements for animal testing.

Box 3: Key concepts cross-fertilizing between translational medicine/biomarkers and the validation of alternatives to animal testing

Focus on human relevance: A major shift involves prioritizing NAMs that directly replicate human biology and re-

[26] https://cipaproject.org
[27] https://fnih.org/our-programs/biomarkers-consortium/

sponses, rather than solely relying on mimicking results from animal models. This leads to more reliable safety and efficacy predictions for humans.

Integrated approaches to testing and assessment: Instead of replacing animal tests one-for-one, the emphasis is on developing intelligent combinations of NAMs (which could include cell-based assays, organ-on-a-chip technology, and computer models). IATAs provide a multi-faceted analysis of a substance's potential effects.

Understanding mechanisms and adverse outcome pathways: There is a push to move from simply identifying hazards to deeply understanding the biological pathways by which a substance causes harm. This allows for better-targeted NAM development and more accurate predictions of risks.

Computational modeling and simulation: Advanced techniques like AI and machine learning are harnessed to analyze vast amounts of data, predict toxicity patterns, and fill in knowledge gaps without the need for animal testing.

Performance standards: Instead of rigid prescriptive protocols, validation should move towards assessing whether a NAM is fit-for-purpose based on its ability to meet specific performance benchmarks relevant to the CoU. This provides flexibility and encourages innovation.

Transparency and reproducibility: Fostering a culture of openness in sharing data, protocols, and method development is essential for building confidence in NAMs.

6 The prospect of adapting a biomarker concept to the NAM validation process

The biomarker concept can bring several valuable aspects to the validation process of NAMs for safety testing:

a) *Mechanistic understanding:* Biomarkers can provide crucial insights into the underlying biological mechanisms of toxicity and offer a more comprehensive understanding of the biological processes involved. This mechanistic understanding can guide the validation process by ensuring that the NAMs are capturing relevant biological events and providing a more accurate assessment of safety. Furthermore, biomarkers can help elucidate complex toxicological cascades, allowing researchers to pinpoint KE in the progression from MIE to AO. This detailed mechanistic knowledge can inform the design and refinement of NAMs, ensuring they target the most relevant biological processes and endpoints.

b) *Predictive power:* Biomarkers that are well established to predict specific toxicological outcomes can significantly enhance the predictive power of NAMs. The inclusion of validated biomarkers in the validation process can strengthen the evidence supporting a NAM's predictive capacity. Combinations of multiple biomarkers can often provide more robust predictions than single markers alone, increasing the sensitivity and specificity of NAMs. The integration of biomarker data with other toxicological information can further enhance the overall predictive performance of these methods.

c) *Translational relevance:* Biomarkers that have been validated in clinical settings or have a strong correlation with human safety outcomes can bridge the gap between preclinical testing and human relevance. Incorporating clinically relevant biomarkers in the validation process can ensure that NAMs are providing meaningful and applicable results for human safety assessment. The use of translational biomarkers can enhance confidence in the validity and relevance of NAMs for predicting human safety. Moreover, translational biomarkers can help address the longstanding challenge of extrapolating from *in vitro* or animal data to human outcomes. By focusing on biomarkers that have demonstrated relevance across species or in human studies, NAMs can provide more reliable predictions of human toxicity, potentially reducing the need for animal testing and improving the efficiency of drug development processes.

d) *Assay performance evaluation:* Biomarkers can serve as performance indicators for evaluating the reliability and reproducibility of NAMs. By measuring specific biomarkers across different laboratories and experimental conditions, the consistency and robustness of NAMs can be assessed. The inclusion of biomarkers in the validation process can help establish performance standards and facilitate the standardization of NAMs. This approach allows for a more objective evaluation of assay performance, as changes in biomarker levels can be quantitatively measured and compared across different experimental setups. Furthermore, using well-characterized biomarkers as benchmarks can help in troubleshooting and optimizing NAMs, ensuring that they consistently produce reliable results across various testing scenarios.

e) *Regulatory acceptance:* The use of well-characterized and clinically relevant biomarkers can facilitate the regulatory acceptance of NAMs. Regulatory agencies are more likely to accept NAMs that incorporate biomarkers with established scientific validity and regulatory precedence. The biomarker qualification process can provide a framework for evaluating the scientific merit and regulatory relevance of NAMs, supporting their acceptance by regulatory authorities. Additionally, biomarkers can serve as a common language between researchers, industry, clinicians and regulators, facilitating communication and understanding of NAM results. As regulatory agencies become more familiar with specific biomarkers and their relationship to toxicity, the incorporation of these biomarkers into NAMs can streamline the regulatory review process and increase confidence in the results obtained from these new methods.

f) *Comparative analysis:* Biomarkers can serve as a common denominator for comparing the performance of different NAMs. By evaluating the ability of NAMs to accurately measure or predict changes in specific biomarkers, the relative effectiveness of different methods can be assessed. This comparative analysis can aid in the selection and prioritization of the most promising NAMs for further validation and implementation.

Tab. 2: Leveraging biomarkers and translational medicine for NAM development

1. Identification of relevant biomarkers
– Identify biomarkers mechanistically linked to toxicological endpoints
– Select biologically relevant and reproducible biomarkers
– Focus on biomarkers strongly associated with toxicological processes
– Design methods to capture key biological events predictive of safety outcomes
2. Incorporation of human-relevant biomarkers
– Emphasize human-relevant models and endpoints
– Use biomarkers validated in human studies or shown to be clinically relevant
– Employ human-derived cells, tissues, or organoids with human-relevant biomarkers
– Improve predictive value by better mimicking human biology
3. Mechanistic understanding
– Gain insights into underlying toxicity mechanisms
– Design methods capturing relevant toxicological processes
– Guide selection of cell types, endpoints, and assay conditions
– Ensure biological relevance and predictivity of alternative methods
4. High-throughput screening
– Integrate biomarkers into high-throughput screening assays
– Rapidly identify potential safety concerns
– Prioritize compounds for further evaluation
– Streamline testing process and reduce animal testing
5. Computational modeling and *in silico* approaches
– Develop models using biomarker data
– Integrate with quantitative structure-activity relationship or physiologically based pharmacokinetic modeling
– Predict safety outcomes without animal testing
– Prioritize compounds and guide targeted *in vitro* assays
6. Validation and qualification of alternative methods
– Use biomarkers as performance indicators
– Assess reliability, reproducibility, and relevance
– Compare biomarker responses to human studies
– Establish predictive value and translational relevance
7. Regulatory acceptance
– Facilitate acceptance through mechanistic understanding
– Incorporate human-relevant biomarkers
– Demonstrate translational value
– Engage early with regulatory authorities and provide robust validation data

Moreover, using biomarkers as benchmarks allows for a more standardized comparison across diverse NAMs, even those based on different technological platforms or biological principles. This standardized approach can facilitate the integration of multiple NAMs into comprehensive testing strategies, leveraging the strengths of each method to provide a more complete toxicological assessment.

g) *Continuous improvement:* Biomarkers allow for the iterative refinement and improvement of NAMs. As new biomarkers are discovered and validated, they can be incorporated into existing NAMs to enhance their predictive power and relevance. The validation process can benefit from the continuous integration of emerging biomarker knowledge, ensuring that NAMs remain up-to-date and aligned with the latest scientific advancements. This dynamic approach to validation allows NAMs to evolve alongside our understanding of toxicology and human biology. Furthermore, the ongoing incorporation of new biomarkers can help address limitations or gaps identified in existing NAMs, continuously improving their performance and applicability. This adaptive validation process ensures that NAMs remain at the forefront of toxicological science, capable of addressing emerging concerns and new classes of compounds.

By leveraging the biomarker concept, the validation process of NAMs can be significantly strengthened, providing a more robust and scientifically sound framework for evaluating the reliability, reproducibility, and relevance of these methods. The incorporation of biomarkers can enhance the mechanistic understanding, predictive power, and translational value of NAMs, ultimately increasing confidence in their ability to accurately assess safety and support regulatory decision-making. This biomarker-centric approach to validation not only improves the scientific basis of NAMs but also facilitates their integration into regulatory frameworks and industry practices, potentially accelerating the adoption of these innovative methods in toxicology and drug development.

By leveraging biomarkers and TM principles, the development of NAMs can be greatly enhanced (Tab. 2). These approaches enable the creation of more predictive, human-relevant, and mechanistically sound testing strategies. By focusing on biologically relevant endpoints, incorporating human-relevant biomarkers, and applying computational modeling techniques, researchers can accelerate the development of NAMs and reduce the reliance on animal testing. Ultimately, this will lead to more efficient and reliable safety assessment, benefiting both human health and animal welfare.

7 Comparison of FDA biomarker qualification to validation of NAMs

The U.S. FDA's Critical Path Initiative[28] emphasized the need for a formal biomarker qualification process to enhance drug development. In response, the FDA's CDER implemented the BQP[29]. This program aims to validate biomarkers for specific CoUs (Amur et al., 2015a). Once a biomarker is qualified through this process, it can be utilized across various drug development programs for its designated CoU without requiring repeated regulatory review. This approach streamlines the integration of validated biomarkers into drug development, potentially accelerating the

[28] https://www.fda.gov/science-research/science-and-research-special-topics/critical-path-initiative

[29] https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/biomarker-qualification-program

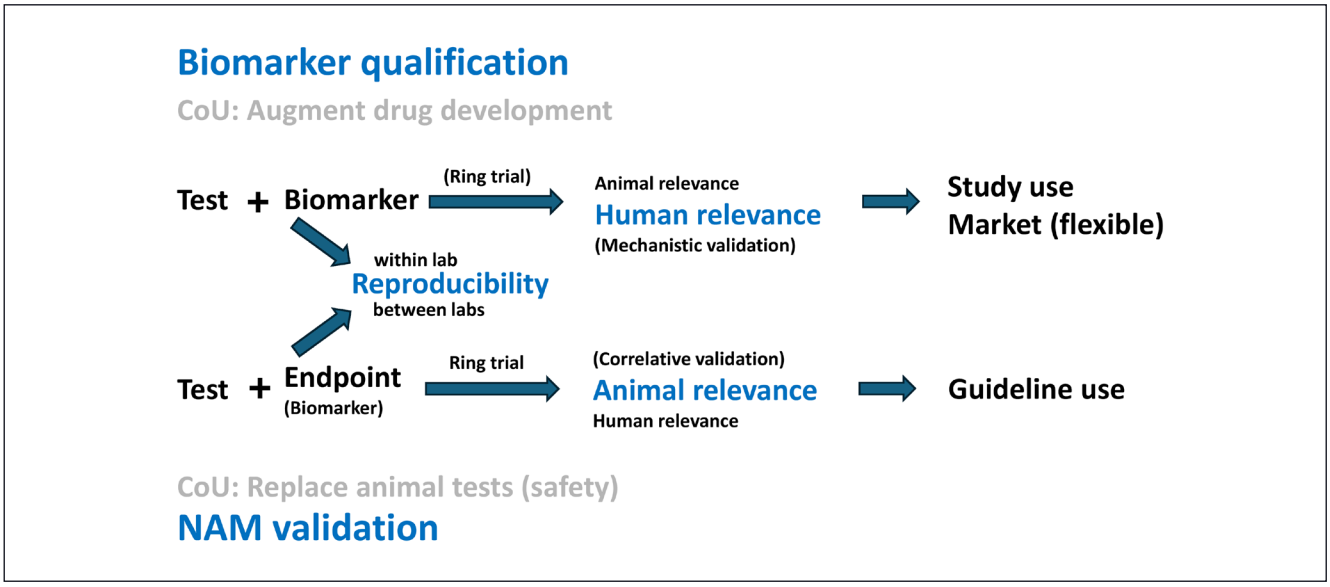

Fig. 2: A simplified comparison of the FDA Biomarker Qualification Process in comparison to traditional NAM validation

overall process and improving efficiency. The FDA BQP (Amur et al., 2015b; Sauer and Porter, 2020) and the validation of NAMs are both rigorous procedures aimed at ensuring the reliability and relevance of new methodologies in biomedical research and drug development. However, they differ in their specific focus, processes, and regulatory contexts. We will summarize the FDA BQP and then compare it in detail to the validation of NAMs.

The FDA BQP:

a) Letter of intent (LOI): Submitters provide an initial description of the biomarker and its CoU.
b) Qualification plan (QP): A detailed plan outlining the studies and analyses to support biomarker qualification.
c) Full qualification package (FQP): Comprehensive submission of all data and analyses supporting the biomarker's use.
d) FDA review: The FDA evaluates the submission and may request additional information.
e) Qualification decision: The FDA decides whether to qualify the biomarker for its proposed use.

As the readers of *ALTEX* will be more familiar with the validation of NAMs, some important terms in the context of biomarkers need to be stressed. Validation of a biomarker focuses on a tool's performance characteristics, while qualification is reserved for a regulatory process. The "evidentiary standards" are then based on the CoU:

a) *Validation:* For biomarkers and clinical outcome assessments (COAs) alike, adequate validation is important for ensuring that a test, tool, or instrument is adequate for its proposed use. Validation of biomarkers requires both analytical and clinical assessment.
b) *Context of use (CoU):* A statement that fully and clearly describes the way the medical product development tool is to be used and the regulated product development and review-related purpose of the use.
c) *Qualification:* A conclusion, based on a formal regulatory process, that within the stated CoU a medical product development tool can be relied upon to have a specific interpretation and application in [medical product development and] regulatory review.

Table 3 compares the biomarker qualification process with the traditional validation of NAMs. In conclusion, while both processes aim to introduce new methodologies into regulatory frameworks, the FDA BQP is more focused on clinical applications and drug development, while the validation of NAMs is broader in scope and more directly aimed at replacing animal use in safety assessment for drug development, food and cosmetic products, and new chemicals. The BQP tends to be more flexible and iterative, while NAM validation often requires more extensive cross-laboratory testing and international consensus.

Figure 2 highlights the differences in emphasis of a biomarker-based NAM validation. Human relevance and mechanistic validation (Hartung et al., 2013b) are in the foreground. There is no need for guidelines as clinical development follows, but there is a market for predictive tests. The within-laboratory reproducibility (necessitating ring trials) is of lesser importance, and proficiency testing of laboratories often suffices. The fact that biomarkers are often offered as commercial kits further supports this.

Interestingly, we are now seeing an increasing number of NAM approaches that do not lend themselves to ring trials as the methods are only offered by a single or very few suppliers. Prominent examples are the high-throughput testing approach of EPA's ToxCast[30] or the Tox-21[31] alliance; similarly, some complex MPS, especially

[30] https://www.epa.gov/comptox-tools/toxicity-forecasting-toxcast
[31] https://tox21.gov

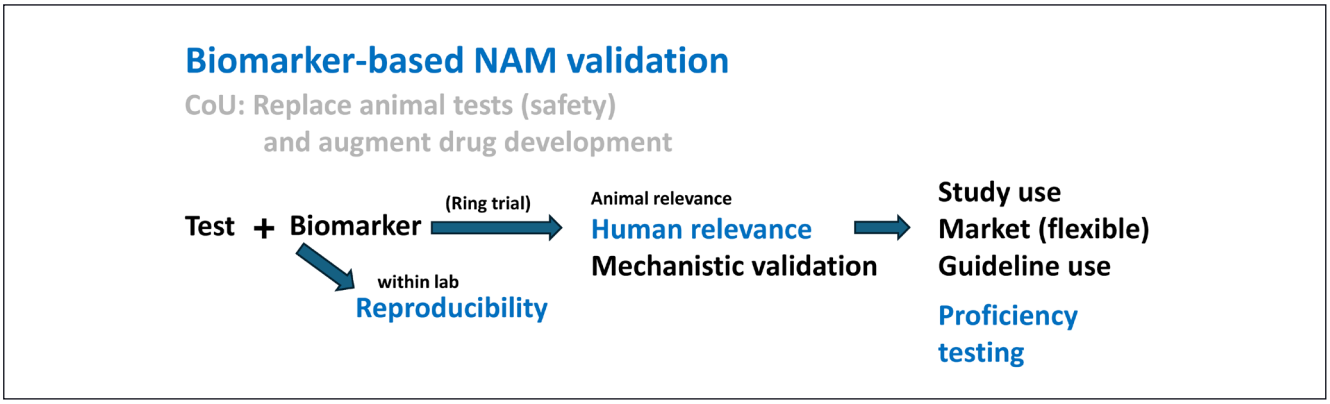

Fig. 3: A simplified presentation on how NAM validation could adapt elements of the FDA Biomarker Qualification Process

Comparison of FDA biomarker qualification to NAM validation

	Biomarker qualification	NAM validation
Purpose and scope	Focuses on validating specific measurable indicators for use in drug development and regulatory decision-making	Aims to replace, reduce, or refine animal testing in toxicology and safety assessment
Regulatory framework	Overseen by the US FDA, with a specific program and guidance (EMA in Europe and PMDA in Japan)	Involves multiple agencies (e.g., EPA, FDA) and international efforts (e.g., ICCVAM, ECVAM, OECD)
Validation process	Three-stage iterative process in the US (LOI, QP, FQP) with the FDA	Typically involves multi-stage validation including pre-validation, validation, and independent peer review
Data requirements	Focuses on human data and clinical relevance	Often compares results to existing animal data or human data where available
Context of use	Specific to the proposed context (e.g., patient selection, dose selection, safety)	Generally aimed at replacing specific animal tests or endpoints
Timeframe	Can take several years, depending on the complexity and available data	Often takes 5-10 years from development to regulatory acceptance
Stakeholder involvement	Primarily involves the submitter and the FDA, with potential input from external experts	Involves a broader range of stakeholders, including academia, industry, regulators, and animal welfare organizations
Acceptance criteria	Focused on demonstrating reliability, sensitivity, and specificity for the proposed context of use	Emphasizes reproducibility, relevance, and ability to predict toxicity in animals
Post-qualification/ validation	Qualified biomarkers are publicly available for use in drug development	Validated methods may be incorporated into regulatory guidelines (e.g., OECD Test Guidelines)
Flexibility	Allows for ongoing refinement and expansion of use through the FDA's BQP	Once validated, changes often require re-validation or adaptation of existing methods
International harmonization	Efforts are ongoing for international harmonization (e.g., through International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)); FDA and EMA have a parallel review option	Strong international cooperation exists (e.g., ICATM)
Endpoint focus	Can cover a wide range of clinical and biological endpoints	Often focused on specific toxicological endpoints (e.g., skin irritation, genotoxicity)

of multiple organs, are often only established in single laboratories. It might be sufficient to assess the within-laboratory performance (Judson et al., 2013) and leave further external validity assessments to a performance standard approach (Hartung et al., 2004) and incorporate proficiency testing in the new laboratories.

Aiming for a synthesis of the two approaches, Figure 3 envisions what NAM validation can learn from the FDA BQP. First and most important is the focus on human relevance, which is enabled by choosing mechanistically relevant biomarkers. De-emphasizing ring trials might be the most controversial part of

the proposal. While ring trials are highly cost- and time-intensive, they are also highly problematic. The effort put into method transfers and training of personnel is not comparable to that of the later adaption by other laboratories. The protocols followed in validation studies are very strict but are loosened later in the guidelines (which may affect reproducibility). On the other hand, a single under-performing laboratory can tank an entire validation study. With proficiency testing, i.e., the laboratory actually carrying out the test later demonstrating that they are capable of doing so, the within-laboratory reproducibility aspect is shifted. The concept of performance standards, introduced with the modular approach to validation (Hartung et al., 2004), which describes which criteria a new test should meet to be considered equivalent to a validated one, could be expanded to proficiency standards to describe what a given laboratory has to show to be deemed able to carry out the test. This encourages also standardization of marketed products to help laboratories meet this criterion. Some of the most successful NAMs have benefitted from this, such as the standardized reagents and training offered for pyrogen tests. The new generation of monocyte activation tests (MAT) now improves standardization by making cryopreserved cells and pre-tested frozen blood available (Schindler et al., 2006; Hartung, 2015, 2021). Arguably, the skin corrosion and skin irritation validations were successful because most of the cell culture work was taken out of the equation: using commercial skin models, most of the variability of cell culture was not part of the assessment. For (regulatory) use or a decision in product development, the result of the ring trial is not really informative but how reliable the result from a proficient laboratory is. By moving the emphasis away from replacing regulatory testing, these methods also enter more easily into investigative toxicology (Beilmann et al., 2019) or green toxicology (Maertens et al., 2014, 2018, 2019, 2024; Maertens and Hartung, 2018).

8 Future perspectives, challenges and opportunities

8.1 Opportunities

There are significant indications of advancement and progress on the organizational level. These include engaging in open dialogue between stakeholders[32] to facilitate regulatory acceptance, establishing international consortia for standardization efforts[33], fostering interdisciplinary collaborations and pre-competitive partnerships[34,35], and increasing investment in novel biomarkers, advanced *in vitro* models, and computational tools[26]. These represent opportunities for the biotech industry to develop and market NAMs. It also represents a key opportunity to align with the biomedical community focusing on pathomechanisms, opening up to the wealth of medical and drug development knowledge as point of reference for validation.

Looking to the future, several trends and technological advances are likely to shape the biomarker field. The integration of multi-omics approaches, combining genomics, proteomics, and metabolomics data for comprehensive toxicity assessment, is expected to provide more holistic insights. Developing integrated biomarker signatures for improved predictivity will become increasingly important. Supporting the creation of advanced *in vitro* models, such as organ-on-a-chip and 3D cell cultures, can improve the physiological relevance of preclinical testing. Advancement of MPS and refining organ-on-a-chip technologies to better mimic human physiology and creating multi-organ systems to assess complex toxicological interactions will enhance the relevance of *in vitro* testing. Investing in computational tools and artificial intelligence for data analysis and modeling can enhance the predictive power of NAMs. Artificial intelligence and machine learning are set to play a crucial role in the future of preclinical safety assessment. Leveraging AI for pattern recognition in large-scale biomarker datasets and developing predictive algorithms for toxicity assessment based on *in vitro* and *in silico* data will greatly enhance our analytical capabilities.

The concept of personalized toxicology is gaining traction. Utilizing patient-derived cells and organoids for individualized risk assessment and incorporating genetic and epigenetic biomarkers for precision toxicology will allow for more tailored safety assessments.

8.2 Challenges

The integration of TM/biomarker methods in preclinical safety assessment faces several key challenges. Ensuring robust validation of biomarkers across diverse populations and conditions affecting susceptibility is crucial for their reliable use. Developing standardized protocols for biomarker measurement and interpretation is necessary to ensure consistency across different research settings. The impact on the NAM validation process of emphasizing biomarker approaches, i.e., the transition to a mechanistic validation, will have to be evaluated. Furthermore, establishing consensus on performance criteria for such biomarker-based NAMs is essential for their widespread adoption.

Scientific challenges also abound in this field. Addressing the complexity of biological systems in *in vitro* and *in silico* models remains a significant hurdle. Developing predictive models that accurately translate *in vitro* findings to *in vivo* outcomes is a key challenge that researchers continue to grapple with.

Developing clear guidelines for the validation and acceptance of alternative methods can streamline their adoption. Implementing a phased approach to introduce alternative methods alongside traditional testing can help build confidence in these new approaches.

Standardization and harmonization efforts present another area of opportunity. Establishing international consortia can promote consistency across studies. Creating shared databases can facilitate cross-study comparisons and meta-analyses. Developing harmonized reporting standards can improve the comparability and reproducibility of results.

32 https://ntp.niehs.nih.gov/go/developers-forums

33 https://www.iqmps.org/

34 https://c-path.org/program/predictive-safety-testing-consortium/

35 https://fnih.org/our-programs/accelerating-medicines-partnership-amp/

Collaborative efforts are crucial for advancing the field. Fostering partnerships between academia, industry, and regulatory agencies can accelerate progress. Establishing pre-competitive collaborations to address common challenges can benefit the entire field. Creating interdisciplinary research teams to tackle complex biological questions can lead to innovative solutions.

Regulatory hurdles present another set of challenges. These regulatory challenges include the need for global harmonization. Overcoming regulatory inertia and resistance to change can be a slow process. Adapting existing regulatory frameworks to accommodate new methodologies requires careful consideration and often lengthy approval processes.

8.3 Limitations and caveats

It is important to realize that the biomarker definition has several components. One of them is that a biomarker must correlate with the process it is a surrogate of. For instance, some AOP in the field of developmental neurotoxicity testing share the KE "disturbed precursor cell differentiation". This may be complex to assess directly but can be quantified in the UKN1 test (Balmer et al., 2012; Rempel et al., 2015; Waldmann et al., 2017; Seidel et al., 2022) by measuring changes in expression of OTX2 and PAX6. These are good biomarkers as their expression correlates under many different circumstances with the differentiation state.

A second component is that the biomarker should ideally correlate well with an AO. This is not necessarily given by the above condition. For instance, in an AOP describing potential DNT effects of neonicotinoids (Loser et al., 2021; Grillberger et al., 2023) "synaptic dysfunction" is a KE. Altered Ca^{2+} signaling may be used as biomarker to assess this KE, but it is not clear at present whether this correlates with an AO. One reason may be that the duration of disturbance plays a role.

A third component, often neglected, is that a biomarker measures a molecular change under certain (defined) conditions in a specified system. In other words, it is context-dependent. A biomarker is not defined as a molecule (or process), because this may qualify as a biomarker under certain conditions, but not under others. A biomarker is also not a certain endpoint, because the same applies as above. For instance, AOP:3 specifies mitochondrial dysfunction as a KE. In some NAMs, induction of ATF4 plus NRF-2 target genes is a good biomarker for this (Delp et al., 2021; Suciu et al., 2023). However, similar gene networks may also be triggered by non-mitochondrial toxicants, and, conversely, in some systems, mitochondrial toxicants may show different gene expression signatures (dependent on organ, energy alternatives, substrate availability, etc.).

In summary, the value and validity of biomarkers can be model-dependent and therefore need to be tested and verified for each model used.

8.4 Prospects and trends

Regulatory evolution is expected to move towards more flexible, science-based frameworks incorporating weight-of-evidence approaches (Linkov et al., 2015) and alternative method data. Global harmonization efforts (Bottini et al., 2007) will be crucial for widespread acceptance of new methodologies and are likely to intensify, with the aim of aligning regulatory requirements across different regions to facilitate global acceptance of alternative methods. Adaptive licensing and progressive authorization approaches may become more common. Implementing staged approval processes that incorporate real-world data and utilizing biomarkers for ongoing safety monitoring post-market approval could revolutionize the drug approval process.

Education and training in new approaches will be essential for successful implementation (Hartung et al., 2009; Daneshian et al., 2011). Updating curricula in toxicology and related fields to incorporate new methodologies and providing ongoing training for regulators and industry professionals on alternative methods will be essential (von Aulock et al., 2022).

9 Conclusion

The exploration of TM/biomarkers in preclinical safety assessment reveals a field poised for significant advancement. It is clear that these approaches hold immense potential for revolutionizing the way we evaluate the safety of drugs and chemicals before they reach human trials or the market. Biomarkers, serving as objective indicators of biological processes, have demonstrated their capacity to provide early, sensitive, and specific insights into potential toxicological and disease-related effects. When combined with the principles of TM, which aims to bridge the gap between laboratory findings and clinical applications, biomarkers offer a powerful toolset for predicting human responses to various compounds.

The potential of these approaches is multifaceted. Firstly, they offer the possibility of more accurate predictions of human responses, addressing the long-standing issue of interspecies differences that often limit the translational value of animal studies. Secondly, they promise the detection of subtle changes at the molecular and cellular levels, potentially identifying safety concerns earlier in the drug development process. This early detection can lead to significant time and cost savings by allowing for the earlier termination of potentially harmful compounds or the refinement of promising ones.

Moreover, the use of TM/biomarker approaches aligns with the growing ethical imperative to reduce reliance on animal testing. As society becomes increasingly aware of animal welfare issues, there is a mounting pressure on the scientific community to develop and adopt alternative methods. The importance of embracing alternative testing methods cannot be overstated. The complexity of human biology and the limitations of animal models in replicating human responses to drugs and chemicals necessitate more sophisticated, human-relevant testing methods. TM/biomarkers, especially when combined with advanced technologies such as organ-on-a-chip models, 3D cell cultures, and *in silico* modeling and/or theoretical constructs (such as AOP), provide a path towards more predictive and reliable preclinical safety assessments.

The journey towards fully realizing this potential should focus on several key areas:

a) Further identification and validation of biomarkers that are predictive of human toxicological responses and their integration

with testing strategies that identify an AO from measurements of the biomarker in a relevant test system.

b) Development of advanced *in vitro* models that recapitulate human physiology and can be used in conjunction with biomarker analyses.

c) Refinement of computational models that can integrate diverse data types to predict human responses.

d) Establishment of standardized protocols and quality control measures to ensure the reliability and reproducibility of these alternative methods.

e) Engagement with regulatory agencies to develop frameworks for the acceptance of data generated using these novel approaches.

In conclusion, TM/biomarkers represent a promising frontier in preclinical safety assessment. They offer the potential to enhance the predictivity, efficiency, and ethical standing of safety testing practices. As we move forward, it is imperative that the scientific community embraces these approaches, investing time, resources, and collaborative efforts into their development and validation. By doing so, we can work towards a future where preclinical safety assessment is more predictive, more efficient, and more aligned with ethical considerations, ultimately leading to safer and more effective therapeutic interventions for human health.

References

Ahmad, A., Imran, M. and Ahsan, H. (2023). Biomarkers as biomedical bioindicators: Approaches and techniques for the detection, analysis, and validation of novel biomarkers of diseases. *Pharmaceutics 15*, 1630. doi:10.3390/pharmaceutics15061630

Amur, S. G., Sanyal, S., Chakravarty, A. G. et al. (2015a). Building a roadmap to biomarker qualification: Challenges and opportunities. *Biomark Med 9*, 1095-1105. doi:10.2217/bmm.15.90

Amur, S., LaVange, L., Zineh, I. et al. (2015b). Biomarker qualification: Toward a multiple stakeholder framework for biomarker development, regulatory acceptance, and utilization. *Clin Pharmacol Ther 98*, 34-46. doi:10.1002/cpt.136

Bakker, E., Hendrikse, N. M., Ehmann, F. et al. (2022). Biomarker qualification at the European medicines agency: A review of biomarker qualification procedures from 2008 to 2020. *Clin Pharmacol Ther 112*, 69-80. doi:10.1002/cpt.2554

Ball, N. M., Bars, R., Botham, P. A. et al. (2022). A framework for chemical safety assessment incorporating new approach methodologies within REACH. *Arch Toxicol 96*, 743-766. doi:10.1007/s00204-021-03215-9

Balmer, N. V., Weng, M. K., Zimmer, B. et al. (2012). Epigenetic changes and disturbed neural development in a human embryonic stem cell-based model relating to the fetal valproate syndrome. *Hum Mol Genet 21*, 4104-4014. doi:10.1093/hmg/dds239

Beilmann, M., Boonen, H., Czich, A. et al. (2019). Optimizing drug discovery by investigative toxicology: Current and future trends. *ALTEX 36*, 3-17. doi:10.14573/altex.1808181

Berman, S. and Siegel, J. (2022). Biomarker qualification at the European medicines agency: A look under the hood. *Clin Pharmacol Ther 112*, 28-30. doi:10.1002/cpt.2609

Blaauboer, B., Boekelheide, K., Clewell, H. J. et al. (2012). The use of biomarkers of toxicity for integrating in vitro hazard estimates into risk assessment for humans. *ALTEX 29*, 411-425. doi:10.14573/altex.2012.4.411

Bleavins, M. R., Carini, C., Jurima-Romet, M. et al. (2010). *Biomarkers in Drug Development: A Handbook of Practice, Application and Strategy*. Hoboken, NJ, USA: Wiley-Blackwell.

Bodaghi, A., Fattahi, N. and Ramazani, A. (2023). Biomarkers: Promising and valuable tools towards diagnosis, prognosis and treatment of Covid-19 and other diseases. *Heliyon 9*, e13323. doi:10.1016/j.heliyon.2023.e13323

Bonventre, J. V., Vaidya, V. S., Schmouder, R. et al. (2010). Next-generation biomarkers for detecting kidney toxicity. *Nat Biotechnol 28*, 436-440. doi:10.1038/nbt0510-436

Bottini, A. A., Amcoff, P. and Hartung, T. (2007). Food for thought … on globalization of alternative methods. *ALTEX 24*, 255-261. doi:10.14573/altex.2007.4.255

Bouhifd, M., Hartung, T., Hogberg, H. T. et al. (2013). Review: Toxicometabolomics. *J Appl Toxicol 33*, 1365-1383. doi:10.1002/jat.2874

Bravo-Merodio, L., Williams, J. A., Gkoutos, G. V. et al. (2019). -Omics biomarker identification pipeline for translational medicine. *J Transl Med 17*, 155. doi:10.1186/s12967-019-1912-5

Brown, K. D., Campbell, C. and Roberts, G. V. (2020). Precision medicine in kidney disease: The patient's view. *Nat Rev Nephrol 16*, 625-627. doi:10.1038/s41581-020-0319-0

Burgdorf, T., Piersma, A., Landsiedel, R. et al. (2019). Workshop on the validation and regulatory acceptance of innovative 3R approaches in regulatory toxicology – Evolution versus revolution. *Toxicol In Vitro 59*, 1-11. doi:10.1016/j.tiv.2019.03.039

Califf, R. M. (2018). Biomarker definitions and their applications. *Exp Biol Med (Maywood) 243*, 213-221. doi:10.1177/1535370217750088

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

Chang, S. Y., Weber, E. J., Sidorenko, V. S. et al. (2017). Human liver-kidney model elucidates the mechanisms of aristolochic acid nephrotoxicity. *JCI Insight 2*, e95978. doi:10.1172/jci.insight.95978

Chen, R., Sanyal, S., Thompson, A. et al. (2018). Evaluating the use of KIM-1 in drug development and research following FDA qualification. *Clin Pharmacol Ther 104*, 1175-1181. doi:10.1002/cpt.1093

Clerbaux, L. A., Amigó, N., Amorim, M. J. et al. (2022). COVID-19 through adverse outcome pathways: Building networks to better understand the disease – 3rd CIAO AOP design workshop. *ALTEX 39*, 322-335. doi:10.14573/altex.2112161

Costa, E., Girotti, S., Pauro, F. et al. (2022). The impact of FDA and EMA regulatory decision-making process on the access to CFTR modulators for the treatment of cystic fibrosis. *Orphanet J Rare Dis 17*, 188. doi:10.1186/s13023-022-02350-5

Cummings, J., Raynaud, F., Jones, L. et al. (2010). Fit-for-purpose biomarker method validation for application in clinical trials of anticancer drugs. *Br J Cancer 103*, 1313-1317. doi:10.1038/sj.

bjc.6605910

Daneshian, M., Akbarsha, M. A., Blaauboer, B. et al. (2011). A framework program for the teaching of alternative methods (replacement, reduction, refinement) to animal experimentation. *ALTEX 28*, 341-352. doi:10.14573/altex.2011.4.341

Das, S., Dey, M. K., Devireddy, R. et al. (2024). Biomarkers in cancer detection, diagnosis, and prognosis. *Sensors 24*, 37. doi:10.3390/s24010037

Davis, K. D., Aghaeepour, N., Ahn, A. H. et al. (2020). Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: Challenges and opportunities. *Nat Rev Neurol 16*, 381-400. doi:10.1038/s41582-020-0362-2

Del Giudice, G., Migliaccio, G., D'Alessandro, N. et al. (2023). Advancing chemical safety assessment through an omics-based characterization of the test system-chemical interaction. *Front Toxicol 5*, 1294780. doi:10.3389/ftox.2023.1294780

Delp, J., Cediel-Ulloa, A., Suciu, I. et al. (2021). Neurotoxicity and underlying cellular changes of 21 mitochondrial respiratory chain inhibitors. *Arch Toxicol 95*, 591-615. doi:10.1007/s00204-020-02970-5

Edington, C. D., Chen, W. L. K., Geishecker, E. et al. (2018). Interconnected microphysiological systems for quantitative biology and pharmacology studies. *Sci Rep 8*, 4530. doi:10.1038/s41598-018-22749-0

El-Achkar, T. M., Eadon, M. T., Kretzler, M. et al. (2024). Precision medicine in nephrology: An integrative framework of multidimensional data in the kidney precision medicine project. *Am J Kidney Dis 83*, 402-410. doi:10.1053/j.ajkd.2023.08.015

Ewart, L., Fabre, K., Chakilam, A. et al. (2017). Navigating tissue chips from development to dissemination: A pharmaceutical industry perspective. *Exp Biol Med 242*, 1579-1585. doi:10.1177/1535370217715441

Ewart, L., Apostolou, A., Briggs, S. A. et al. (2022). Performance assessment and economic analysis of a human liver-chip for predictive toxicology. *Commun Med (Lond) 2*, 154. doi:10.1038/s43856-022-00209-1

FDA-NIH Biomarker Working Group (2016-). Monitoring biomarker. 2016 Dec 22 [Updated 2021 Jan 25]. In *BEST (Biomarkers, EndpointS, and other Tools)*. Silver Spring, MD, USA: Food and Drug Administration (US) Co-published by Bethesda, MD, USA: National Institutes of Health (US). https://www.ncbi.nlm.nih.gov/books/nbk402282/

Grillberger, K., Cöllen, E., Trivisani, C. I. et al. (2023). Structural insights into neonicotinoids and N-unsubstituted metabolites on human nAChRs by molecular docking, dynamics simulations, and calcium imaging. *Int J Mol Sci 24*, 13170. doi:10.3390/ijms241713170

Gromova, M., Vaggelas, A., Dallmann, G. et al. (2020). Biomarkers: Opportunities and challenges for drug development in the current regulatory landscape. *Biomark Insights 15*, 1177271920974652. doi:10.1177/1177271920974652

Guo, L., Coyle, L., Abrams, R. M. C. et al. (2013). Refining the human iPSC-cardiomyocyte arrhythmic risk assessment model. *Toxicol Sci 136*, 581-594. doi:10.1093/toxsci/kft205

Hargrove-Grimes, P., Low, L. A. and Tagle, D. A. (2021). Microphysiological systems: What it takes for community adoption. *Exp Biol Med (Maywood) 246*, 1435-1446. doi:10.1177/15353702211008872

Hargrove-Grimes, P., Low, L. A. and Tagle, D. A. (2022). Microphysiological systems: Stakeholder challenges to adoption in drug development. *Cells Tissues Organs 211*, 269-281. doi:10.1159/000517422

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. and Leist, M. (2008). Food for thought … on the evolution of toxicology and phasing out of animal testing. *ALTEX 25*, 91-96. doi:10.14573/altex.2008.2.91

Hartung, T., Blaauboer, B. and Leist, M. (2009). Food for thought … on education in alternative methods in toxicology. *ALTEX 26*, 255-263. doi:10.14573/altex.2009.4.255

Hartung, T. (2013). Look back in anger – What clinical studies tell us about preclinical work. *ALTEX 30*, 275-291. doi:10.14573/altex.2013.3.275

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

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

Hartung T. (2015). The human whole blood pyrogen test – Lessons learned in twenty years. *ALTEX 32*, 79-100. doi:10.14573/altex.1503241

Hartung, T. (2017). Utility of the adverse outcome pathway concept in drug development. *Expert Opin Drug Metab Toxicol 13*, 1-3. doi:10.1080/17425255.2017.1246535

Hartung, T. (2021). Pyrogen testing revisited on occasion of the 25th anniversary of the whole blood test. *ALTEX 38*, 3-19. doi:10.14573/altex.2101051

Hartung, T. (2023a). A call for a human exposome project. *ALTEX 40*, 4-33. doi:10.14573/altex.2301061

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

Hartung, T. (2024a). The (misleading) role of animal models in drug development. *Front Drug Discov 4*, 1355044. doi:10.3389/fddsv.2024.1355044

Hartung, T. (2024b). The validation of regulatory test methods – Conceptual, ethical, and philosophical foundations. *ALTEX 41*, 525-544. doi:10.14573/altex.2409271

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

Hendrikse, N. M., Llinares Garcia, J., Vetter, T. et al. (2022). Biomarkers in medicines development – From discovery to regulatory qualification and beyond. *Front Med 9*, 878942. doi:10.3389/fmed.2022.878942

Hewitt, N. J., Bühring, K. U., Dasenbrock, J. et al. (2001). Studies comparing in vivo:in vitro metabolism of three pharmaceutical compounds in rat, dog, monkey, and human using cryopreserved

hepatocytes, microsomes, and collagen gel immobilized hepatocyte cultures. *Drug Metab Dispos 29*, 1042-1050.

Hood, L., Heath, J. R., Phelps, M. E. et al. (2004). Systems biology and new technologies enable predictive and preventative medicine. *Science 306*, 640-643. doi:10.1126/science.1104635

Horien, C. (2017). Biomarkers, translational medicine, and drug development: An interview with Chirag R. Parikh, MD, PhD. *Yale J Biol Med 90*, 153-156.

Institute of Medicine (US) Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease (2010). 4 case studies. In C. M. Micheel and J. R. Ball (eds), *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease*. Washington, DC, USA: National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/nbk220298/

Judson, R., Kavlock, R., Martin, M. et al. (2013). Perspectives on validation of high-throughput pathway-based assays supporting the 21st century toxicity testing vision. *ALTEX 30*, 51-66. doi:10.14573/altex.2013.1.051

Kimmel, P. L., Jefferson, N., Norton, J. M. (2019). How community engagement is enhancing NIDDK research. *Clin J Am Soc Nephrol 14*, 768-770. doi:10.2215/cjn.14591218

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

Kopec, A. K., Yokokawa, R., Khan, N. et al. (2021). Microphysiological systems in early stage drug development: Perspectives on current applications and future impact. *J Toxicol Sci 46*, 99-114. doi:10.2131/jts.46.99

Krewski, D., Westphal, M., Andersen, M. E. et al. (2014). A framework for the next generation of risk science. *Environ Health Perspect 122*, 796-805. doi:10.1289/ehp.1307260

Kumar, C. and van Gool, A. J. (2013). Introduction: Biomarkers in translational and personalized medicine. In P. Horvatovich and R. Bischoff (eds), *Comprehensive Biomarker Discovery and Validation for Clinical Application* (Chapter 1, 3-39). The Royal Society of Chemistry.

Lee, J. W., Devanarayan, V., Barrett, Y. C. et al. (2006). Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm Res 23*, 312-328. doi:10.1007/s11095-005-9045-3

Leist, M., Hartung, T. and Nicotera, P. (2008). The dawning of a new age of toxicology. *ALTEX 25*, 103-114. doi:10.14573/altex.2008.2.103

Leist, M., Ghallab, A., Graepel, R. et al. (2017). Adverse outcome pathways: Opportunities, limitations and open questions. *Arch Toxicol 31*, 221-229. doi:10.1007/s00204-017-2045-3

Lin, Z. and Chou, W. C. (2022). Machine learning and artificial intelligence in toxicological sciences. *Toxicol Sci 189*, 7-19. doi:10.1093/toxsci/kfac075

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

Loser, D., Hinojosa, M. G., Blum, J. et al. (2021). Functional alterations by a subgroup of neonicotinoid pesticides in human dopaminergic neurons. *Arch Toxicol 95*, 2081-2107. doi:10.1007/s00204-021-03031-1

Luciano, J. S., Andersson, B., Batchelor, C. et al. (2011). The translational medicine ontology and knowledge base: Driving personalized medicine by bridging the gap between bench and bedside. *J Biomed Semant 2, Suppl 2*, S1. doi:10.1186/2041-1480-2-s2-s1

Maertens, A., Anastas, N., Spencer, P. J. et al. (2014). Green toxicology. *ALTEX 31*, 243-249. doi:10.14573/altex.1406181

Maertens, A. and Hartung, T. (2018). Green toxicology – Know early about and avoid toxic product liabilities. *Toxicol Sci 161*, 285-289. doi:10.1093/toxsci/kfx243

Maertens, A., Golden, E. and Hartung, T. (2019). Avoiding regrettable substitutions: Green toxicology for sustainable chemistry. *ACS Sustain Chem Eng 9*, 7749-7758. doi:10.1021/acssuschemeng.0c09435

Maertens, A., Luechtefeld, T. and Hartung, T. (2024). Alternative methods go green! Green toxicology as a sustainable approach for assessing chemical safety and designing safer chemicals. *ALTEX 41*, 3-19. doi:10.14573/altex.2312291

Magurany, K. A., Chang, X., Clewell, R. et al. (2023). A pragmatic framework for the application of new approach methodologies in one health toxicological risk assessment. *Toxicol Sci 192*, 155-177. doi:10.1093/toxsci/kfad012

Mahony, C. (2019). Building confidence in non-animal methods: Practical examples of collaboration between regulators, researchers and industry. *Comput Toxicol 10*, 78-80. doi:10.1016/j.comtox.2019.01.003

Malik, M., Yang, Y., Fathi, P. et al. (2021). Critical considerations for the design of multi-organ microphysiological systems (MPS). *Front Cell Dev Biol 9*, 721338. doi:10.3389/fcell.2021.721338

Mankoff, S. P., Brander, C., Ferrone, S. et al. (2004). Lost in translation: Obstacles to translational medicine. *J Transl Med 2*, 14. doi:10.1186/1479-5876-2-14

Mansouri, M., Lam, J. and Sung, K. E. (2024). Progress in developing microphysiological systems for biological product assessment. *Lab Chip 24*, 1293-1306. doi:10.1039/d3lc00876b

Marciano, L. P. A., Kleinstreuer, N., Chang, X. et al. (2024). A novel approach to triazole fungicides risk characterization: Bridging human biomonitoring and computational toxicology. *Sci Total Environ 953*, 176003. doi:10.1016/j.scitotenv.2024.176003

Marx, U., Andersson, T. B., Bahinski, A. et al. (2016). Biology-inspired microphysiological system approaches to solve the prediction dilemma of substance testing using animals. *ALTEX 33*, 272-321. doi:10.14573/altex.1603161

Marx, U., Akabane, T., Andersson, T. B. et al. (2020). Biology-inspired microphysiological systems to advance medicines for patient benefit and animal welfare. *ALTEX 37*, 364-394. doi:10.14573/altex.2001241

Mattes, W. and Walker, E. (2009). Translational toxicology and the work of the predictive safety testing consortium. *Clin Pharmacol Ther 85*, 327-330. doi:10.1038/clpt.2008.270

Mattes, W. B., Gribble Walker, E., Abadie, E. et al. (2010). Research at the interface of industry, academia and regulatory science. *Nat Biotechnol 28*, 432-433. doi:10.1038/nbt0510-432

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

Monticello, T. M., Jones, T. W., Dambach, D. M. et al. (2017). Current nonclinical testing paradigm enables safe entry to first-In-human clinical trials: The IQ consortium nonclinical to clinical translational database. *Toxicol Appl Pharmacol 334*, 100-109. doi:10.1016/j.taap.2017.09.006

Murad, N. and Melamud, E. (2022). Global patterns of prognostic biomarkers across disease space. *Sci Rep 12*, 21893. doi:10.1038/s41598-022-25209-y

Nendza, M. and Ahlers, J. (2022). Aquatic toxicity integrated testing and assessment strategies (ITS) for difficult substances: Case study with thiochemicals. *Environ Sci Eur 34*, 17. doi:10.1186/s12302-022-00591-6

Nguyen, V. V. T., Gkouzioti, V., Maass, C. et al. (2023). A systematic review of kidney-on-a-chip-based models to study human renal (patho-)physiology. *Dis Model Mech 16*, dmm050113. doi:10.1242/dmm.050113

OECD (2023). Guideline No. 497: Defined Approaches on Skin Sensitisation. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/b92879a4-en

Olson, H., Betton, G., Robinson, D. et al. (2000). Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol 32*, 56-67. doi:10.1006/rtph.2000.1399

Owen, M., Bose, N., Nisenbaum, L. et al. (2023). The critical role of biomarkers for drug development targeting the biology of aging. *J Prev Alzheimers Dis 10*, 729-742. doi:10.14283/jpad.2023.111

Ozer, J. S., Dieterle, F., Troth, S. et al. (2010). A panel of urinary biomarkers to monitor reversibility of renal injury and a serum marker with improved potential to assess renal function. *Nat Biotechnol 28*, 486-494. doi:10.1038/nbt.1627

Pamies, D., Leist, M., Coecke, S. et al. (2022). Guidance document on Good Cell and Tissue Culture Practice 2.0 (GCCP 2.0). *ALTEX 39*, 30-70. doi:10.14573/altex.2111011

Pamies, D., Ekert, J., Zurich, M. G. et al. (2024). Recommendations on fit-for-purpose criteria to establish quality management for microphysiological systems and for monitoring their reproducibility. *Stem Cell Reports 19*, 604-617. doi:10.1016/j.stemcr.2024.03.009

Pang, L., Cai, C., Aggarwal, P. et al. (2024). Predicting oncology drug-induced cardiotoxicity with donor-specific iPSC-CMs – A proof-of-concept study with doxorubicin. *Toxicol Sci 200*, 79-94. doi:10.1093/toxsci/kfae041

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

Ramirez, T., Daneshian, M., Kamp, H. et al. (2013). Metabolomics in toxicology and preclinical research. *ALTEX 30*, 209-225. doi:10.14573/altex.2013.2.209

Ratner, M. (2017). FDA deems in vitro data on mutations sufficient to expand cystic fibrosis drug label. *Nat Biotechnol 35*, 606-606. doi:10.1038/nbt0717-606

Rempel, E., Hoelting, L., Waldmann, T. et al. (2015). A transcriptome-based classifier to identify developmental toxicants by stem cell testing: Design, validation and optimization for histone deacetylase inhibitors. *Arch Toxicol 89*, 1599-1618. doi:10.1007/s00204-015-1573-y

Robinson, W. H., Lindstrom, T. M., Cheung, R. K. et al. (2013). Mechanistic biomarkers for clinical decision making in rheumatic diseases. *Nat Rev Rheumatol 9*, 267-276. doi:10.1038/nrrheum.2013.14

Roth, A. and MPS-WS Berlin 2019 (2021). Human microphysiological systems for drug development. *Science 373*, 1304-1306. doi:10.1126/science.abc3734

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

Sasseville, V. G., Mansfield, K. G. and Brees, D. J. (2014). Safety biomarkers in preclinical development: Translational potential. *Vet Pathol 51*, 281-291. doi:10.1177/0300985813505117

Sauer, J. M. and Porter, A. C. (2018). Preclinical biomarker qualification. *Exp Biol Med (Maywood) 243*, 222-227. doi:10.1177/1535370217743949

Sauer, J. M. and Porter, A. C. (2020). The regulatory acceptance of translational safety biomarkers. *Curr Opin Toxicol 23-24*, 80-86. doi:10.1016/j.cotox.2020.06.001

Schindler, S., Spreitzer, I., Löschner, B. et al. (2006). International validation of pyrogen tests based on cryopreserved human primary blood cells. *J Immunol Methods 316*, 42-51. doi:10.1016/j.jim.2006.07.023

Schmeisser, S., Miccoli, A., von Bergen, M. et al. (2023). New approach methodologies in human regulatory toxicology - Not if, but how and when!. *Environ Int 178*, 108082. doi:10.1016/j.envint.2023.108082

Schomaker, S., Ramaiah, S., Khan, N. et al. (2019). Safety biomarker applications in drug development. *J Toxicol Sci 44*, 225-235. doi:10.2131/jts.44.225

Schwartz, M. P., Hou, Z., Propson, N. E. et al. (2015). Human pluripotent stem cell-derived neural constructs for predicting neural toxicity. *Proc Natl Acad Sci U S A 112*, 12516-12521. doi:10.1073/pnas.1516645112

Seidel, F., Cherianidou, A., Kappenberg, F. et al. (2022). High accuracy classification of developmental toxicants by in vitro tests of human neuroepithelial and cardiomyoblast differentiation. *Cells 11*, 3404. doi:10.3390/cells11213404

Serelli-Lee, V., Ito, K., Koibuchi, A. et al. (2022). A state-of-the-art roadmap for biomarker-driven drug development in the era of personalized therapies. *J Pers Med 12*, 669. doi:10.3390/jpm12050669

Sillé, F. and Hartung, T. (2024). Metabolomics in preclinical drug safety assessment: Current status and future trends. *Metabolites 14*, 98. doi:10.3390/metabo14020098

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

Skolariki, K., Exarchos, T. P. and Vlamos, P. (2023). Computa-

tional models for biomarker discovery. *Adv Exp Med Biol 1424*, 289-295. doi:10.1007/978-3-031-31982-2_33

Song, X. and Dobbin, K. K. (2022). Evaluating biomarkers for treatment selection from reproducibility studies. *Biostatistics 23*, 173-188. doi:10.1093/biostatistics/kxaa018

Stucki, A. O., Barton-Maclaren, T. S., Bhuller, Y. et al. (2022). Use of new approach methodologies (NAMs) to meet regulatory requirements for the assessment of industrial chemicals and pesticides for effects on human health. *Front Toxicol 4*, 964553. doi:10.3389/ftox.2022.964553

Suciu, I., Delp, J., Gutbier, S. et al. (2023). Definition of the neurotoxicity-associated metabolic signature triggered by berberine and other respiratory chain inhibitors. *Antioxidants 13*, 49. doi:10.3390/antiox13010049

Sung, N. S., Crowley, W. F. Jr., Genel, M. et al. (2003). Central challenges facing the national clinical research enterprise. *JAMA 289*, 1278-1287. doi:10.1001/jama.289.10.1278

Tagle, D. A. (2019). The NIH microphysiological systems program: Developing in vitro tools for safety and efficacy in drug development. *Curr Opin Pharmacol 48*, 146-154, doi:10.1016/j.coph.2019.09.007

Takeda, M., Miyagawa, S., Fukushima, S. et al. (2018). Development of in vitro drug-induced cardiotoxicity assay by using three-dimensional cardiac tissues derived from human induced pluripotent stem cells. *Tissue Eng Part C Methods 24*, 56-67. doi:10.1089/ten.tec.2017.0247

Tonoyan, L. and Siraki, A. G. (2024). Machine learning in toxicological sciences: Opportunities for assessing drug toxicity. *Front Drug Discov 4*, 1336025. doi:10.3389/fddsv.2024.1336025

Tran, T. T. V., Wibowo, A. S., Tayara, H. et al. (2023). Artificial intelligence in drug toxicity prediction: Recent advances, challenges, and future perspectives. *J Chem Inf Model 63*, 2628-2643. doi:10.1021/acs.jcim.3c00200

Troth, S. P., Vlasakova, K., Amur, S. et al. (2019). Translational safety biomarkers of kidney injury. *Semin Nephrol 39*, 202-214. doi:10.1016/j.semnephrol.2018.12.008

Tsaioun, K., Blaauboer, B. J. and Hartung, T. (2016). Evidence-based absorption, distribution, metabolism, excretion and toxicity (ADMET) and the role of alternative methods. *ALTEX 33*, 343-358. doi:10.14573/altex.1610101

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

Vlasakova, K., Bourque, J., Bailey, W. J. et al. (2022). Universal accessible biomarkers of drug-induced tissue injury and systemic inflammation in rat: Performance assessment of TIMP-1, A2M, AGP, NGAL, and albumin. *Toxicol Sci 187*, 219-233. doi:10.1093/toxsci/kfac030

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

Wagner, P. D. and Srivastava, S. (2012). New paradigms in translational science research in cancer biomarkers. *Transl Res 159*, 343-353. doi:10.1016/j.trsl.2012.01.015

Waldmann, T., Grinberg, M., König, A. et al. (2017). Stem cell transcriptome responses and corresponding biomarkers that indicate the transition from adaptive responses to cytotoxicity. *Chem Res Toxicol 30*, 905-922. doi:10.1021/acs.chemrestox.6b00259

Wang, X. and Ward, P. A. (2012). Opportunities and challenges of disease biomarkers: A new section in the journal of translational medicine. *J Transl Med 10*, 240. doi:10.1186/1479-5876-10-240

Weaver, J. L., Wu, W., Hyland, P. L. et al. (2022). Expanding approved patient populations for rare disease treatment using in vitro data. *Clin Pharmacol Ther 112*, 58-61. doi:10.1002/cpt.2414

Wehling, M. (2015). *Principles of Translational Science in Medicine: From Bench to Bedside* (2nd edition). Elsevier Science.

Wilmer, M. J., Ng, C. P., Lanz, H. L. et al. (2016). Kidney-on-a-chip technology for drug-induced nephrotoxicity screening. *Trends Biotechnol 34*, 156-170. doi:10.1016/j.tibtech.2015.11.001

Yang, X., Ribeiro, A. J. S., Pang, L. et al. (2022). Use of human iPSC-CMs in nonclinical regulatory studies for cardiac safety assessment. *Toxicol Sci 190*, 117-126. doi:10.1093/toxsci/kfac095

Disclaimer

The spectrum of views expressed in this article are those of the contributing authors and do not necessarily reflect those of their institution of employment.

Conflict of interest

The authors declare no conflict of interest.

Data availability

No datasets were generated for this article.

Acknowledgements

The valuable discussions with Dr Suzie Fitzpatrick, Dr William Mattes, and Dr Shaun MacMahon from the Food and Drug Administration (FDA), Center for Food Safety and Applied Nutrition (CFSAN), are very much appreciated. We gratefully acknowledge funding by EFSA and the Doerenkamp-Zbinden Foundation. The work was also facilitated by grants from the BMBF (161L0243B, 016LW0146K) and the *Land Baden-Württemberg* (NAM_ACCEPT, INVITE). Funding was also received from the European Union's Horizon 2020 research and innovation program under grant agreements No. 964537 (RISK-HUNT*3R*), No. 964518 (ToxFree), No. 963845 (ONTOX) and No. 101057014 (PARC), and by the IHI2 project Vict3R.