EUROPEAN
COMMISSION

Brussels, 1.6.2026
SWD(2026) 144 final

COMMISSION STAFF WORKING DOCUMENT
Accompanying the document

Communication from the Commission

Roadmap towards phasing out animal testing for chemical safety assessments

{C(2026) 3497 final}

Contents

Contents 1
List of abbreviations 3
Introduction 6
1. Developing a roadmap – a collective effort 7
1.1. Roadmap development through inter-service and inter-agency collaboration 7
1.2. Building a community through dedicated conferences and workshops 7
1.3. Listening to stakeholders – gathering evidence through consultations 7
1.4. Understanding change – stakeholder perspectives 8
2. Identifying opportunities towards phasing out animal use 9
2.1. Recommendation of actions for human health and environmental assessments 9
2.2. Preparing for the transition to animal-free chemical safety assessment 16
2.3. Overarching opportunities across Human Health and Environmental Safety 53
3. Recommendations and actions for environmental safety assessments 61
3.1. Acute fish toxicity 61
3.2. Aquatic and terrestrial bioaccumulation 65
3.3. Fish chronic toxicity 78
3.4. Endocrine disruption for environmental safety assessments 90
3.5. Birds and mammals 100
3.6. Ecological thresholds of toxicological concern and species sensitivity distributions 107
4. Recommendations and actions for human health assessments 109
4.1. Acute toxicity (human health assessment) 109
4.2. Genotoxicity 116
4.3. Carcinogenicity 125
4.4. Repeated dose toxicity 135
4.5. Developmental neurotoxicity - the DNT in vitro battery 163
4.6. Developmental and reproductive toxicity (DART) assessments 175
4.7. Endocrine Disrupting Chemicals - Use of New Approach Methodologies for their assessment within European Union regulatory frameworks: Recommendations from the 2024 EPAA Partners' Forum 186
5. Working together to design and implement non-animal approaches in regulatory testing 196
5.1. Role and composition of the roadmap steering team 196
5.2. Roadmap steering team 196
5.3. Engaging Member-State and stakeholder experts in agency collaborative structures 198
5.4. EU Agency Network (EUAN) – working group on animal-free approaches 199
5.5. Translating innovative methods into regulatory applications – validation, standardisation, qualification 201

5.6. Safe spaces and regulatory exploration spaces for animal-free methodologies 207
6. Indicators – managing change by measuring it.....212
Annexes 230
Annex I – Chemical safety assessments under following legislations are in scope of the roadmap 230
Annex II - Ongoing agency activities (ECHA, EFSA and EMA) that support the phasing out of animal testing for chemical safety assessments 234
Annex III - Engaging with the general public 244

List of abbreviations

3Rs – replacement, reduction and refinement
3Rs WP – European Medicines Agency working party on the 3Rs
ADME – absorption, distribution, metabolism and excretion
ACR – acute-to-chronic ratio
AFSA – animal-free safety assessment
AI – artificial intelligence
AOP – adverse outcome pathway
APCRA – Accelerating the Pace of Chemical Risk Assessment
ASPA – ASPIS-initiated alternative safety profiling approach
ASPIS – animal-free safety assessment of chemicals: project cluster for implementation of novel strategies
BCF – bioconcentration factor
BfR – Bundesinstitut für Risikobewertung (German Federal Institute for Risk Assessment)
BMF – biomagnification factor
BPR - Biocidal Products Regulation
CEN – European Committee for Standardization
CF – conceptual framework
CHMP – Committee for Medicinal Products for Human Use
C&L – classification and labelling
CLP – classification, labelling and packaging
CRO – contract research organisation
CSR – corporate social responsibility
DA – defined approach
DART – developmental and reproductive toxicity
DNT – developmental neurotoxicity
EBW – exposure-based waiving
EC – European Commission
ECHA – European Chemicals Agency
ECI – European Citizens' Initiative
ecoTTC – ecological threshold of toxicological concern
ED – endocrine disruption
EFSA – European Food Safety Authority
EMA – European Medicines Agency
EPAA – European Partnership for Alternative Approaches to Animal Testing
ERA – European Research Area
ESA – environmental safety assessment
ESEC – European specialised expert community
EU – European Union
EU NETVAL – EU network of validation laboratories
EURL ECVAM – EU Reference Laboratory for Alternatives to Animal Testing
FET – fish embryo test
GD – guidance document
GHS – globally harmonised system

GLP – good laboratory practice
GMO – genetically modified organism
HSI – Humane Society International
IATA – integrated approaches to testing and assessment
ICCS – International Collaboration on Cosmetics Safety
ICCVAM – Interagency Coordinating Committee on the Validation of Alternative Methods
ICH – International Council on Harmonisation
IMI – Innovative Medicines Initiative
IOC – ionisable organic compounds
ISG – inter-service group of Commission services and the three EU-Agencies ECHA, EFSA and EMA
ISO – International Organization for Standardization
IVIVE – in vitro-in vivo extrapolation
JRC – Joint Research Centre
KE – key event
KER – key event relationships
LD50 – median lethal dose (lethal dose 50%)
LERU – League of European Research Universities
LoE – lines of evidence
MAD – mutual acceptance of data
MIE – molecular initiating event
ML – machine learning
MoA – mode of action
MPS – micro-physiological systems
NAM – new approach methodology
NGO – non-governmental organisation
NGRA – next-generation risk assessment
NIVA – Norsk institutt for vannforskning (Norwegian Institute for Water Research)
NOAEL – no observed adverse effect level
NoG – notes of guidance
OECD – Organisation for Economic Co-operation and Development
OoC – organ-on-chip
OSOA – one substance, one assessment
PARC – Partnership for the Assessment of Risks from Chemicals
PBT/PMT – persistent, bioaccumulative and toxic / persistent, mobile and toxic
PBTK – physiologically-based toxicokinetic
PBK – physiologically-based kinetic
PEC – predicted environmental concentration
PoD – point of departure
PNEC – predicted no-effect concentration
qAOP – quantitative adverse outcome pathway
QIVIVE – quantitative in vitro-in vivo extrapolation
OPPTS – Office of Prevention, Pesticides, and Toxic Substances
QSAR – quantitative structure-activity relationship

REACH – registration, evaluation, authorisation and restriction of chemicals
RES – regulatory exploration space
SCCS – Scientific Committee on Consumer Safety
SIR – standard information requirement
SME – small and medium-sized enterprise
SSbD – safe and sustainable by design
SSD – species sensitivity distribution
TD – toxicodynamics
TGs – test guidelines
TK – toxicokinetics
TTC – threshold of toxicological concern
UVCB – unknown or variable composition, complex reaction products, or biological materials
UN – United Nations
US – United States
WHO – World Health Organization
WoE – weight of evidence

Introduction

In 2023, responding to the European Citizens' Initiative (ECI) 'Save Cruelty-Free Cosmetics – Commit to a Europe Without Animal Testing', the Commission committed to developing a roadmap towards phasing out animal testing for chemical safety assessments. In response to both this commitment and repeated calls from the European Parliament for a strategic plan, the Commission is now introducing a roadmap in the Communication corresponding to this staff working document (SWD). This SWD contains more detailed information supporting each section of the Communication.

The roadmap applies to all pieces of EU legislation that require animal testing for chemical safety assessments. A total of 15 different legislative areas have been identified that fall under the roadmap's scope. Annex I provides an overview of these legislative areas.

The goal of this roadmap – to phase out animal testing for chemical safety assessments – cannot be achieved solely through actions of the Commission. Therefore, numerous stakeholders have contributed to the roadmap's development and have committed to undertake many of the individual actions described in this SWD.

This SWD outlines:

- opportunities and concrete actions per toxicological endpoint across the main areas of human health and environmental safety assessment, identified by the working groups for the safety assessment of human health and the environment during roadmap development, as summarised under 2. Identifying opportunities towards phasing out animal use;

- the proposed organisational structures to support the roadmap's implementation, existing safe-space structures and processes, and agency activities that support the phasing out of animal testing for chemical safety assessments, as summarised under 5. Working together to design and implement non-animal approaches in regulatory testing;

- reflections on possible ways for engaging with the public to ensure broad acceptance of the proposed revised approach to how chemical safety assessments are conducted, as summarised in Annex III: Engaging with the general public;

- possible indicators for monitoring the progress of roadmap implementation, as summarised under 6. Indicators- Managing change by measuring it;

- ongoing agency activities (the European Chemicals Agency (ECHA), the European Food Safety Authority (EFSA) and the European Medicines Agency (EMA)) that support the phasing out of animal testing for chemical safety assessments (Annex II).

1. Developing a roadmap – a collective effort

1.1. Roadmap development through inter-service and inter-agency collaboration

The Commission set up an inter-service steering group consisting of relevant services [1] as well as ECHA, EFSA and the EMA to guide the roadmap's development. Three working groups were also created to develop the specific recommendations described in the respective chapters of this SWD for safety assessments of human health (section 4) and environmental safety (section 5) and for change management. The individual chapters in sections 4 and 5 on selected toxicology endpoints were developed by the respective working groups, where stakeholders contributed to drafting specific chapters.

1.2. Building a community through dedicated conferences and workshops

To ensure that stakeholders were involved in the design of the roadmap, the Commission organised three workshops on this topic.

The first workshop on 11-12 December 2023 sought to identify the major challenges in moving towards non-animal chemical safety assessments and the ways of achieving this goal.

A second workshop organised on 25 October 2024 collected input from various stakeholders on how they addressed the transition from animal-based testing to innovative, reliable non-animal alternatives. Participants reported on progress in developing the roadmap and discussed actions and milestones for driving forward this transformation of chemical safety assessments.

The third workshop on 16 -17 June 2025 presented a first draft of the roadmap's content and collected feedback from stakeholders on the potential actions and milestones.

A three-day conference was co-organised by the Commission and the European Partnership for Alternative Approaches to Animal Testing (EPAA) in March 2025 to develop proposals for short-, medium- and long-term actions for potential inclusion in the roadmap.

Several other events were also organised by stakeholders to provide information and exchange views on the development of the roadmap.

All these events gathered a large number of experts from various stakeholder groups, bringing highly valuable insights to the roadmap's development.[2]

1.3. Listening to stakeholders – gathering evidence through consultations

To ensure that a broad range of stakeholders were involved in the preparation of this roadmap, several consultation activities were carried out in addition to the events described above.

First, a call for evidence was organised according to the Better Regulation guidelines[3] via the 'Have your Say'[4] website, with a four-week consultation period in September – October 2024. As a result, 90

[1] DG GROW, ENV, JRC, RTD, SANTE and SG.
[2] https://single-market-economy.ec.europa.eu/sectors/chemicals/reach/roadmap-towards-phasing-out-animal-testing/events-and-outreach-activities_en.
[3] https://commission.europa.eu/law/law-making-process/better-regulation/better-regulation-guidelines-and-toolbox_en.
[4] https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14281-Animal-testing-in-chemical-safety-assessments-Commission-roadmap-to-phase-it-out_en.

responses were submitted and are available on the 'Have your Say' site. A summary of the received comments is available in a synopsis report [5].

In addition, two targeted online stakeholder surveys and expert interviews were organised to collect feedback on more specific issues pertinent to the roadmap's development. Feedback was received from a broad set of stakeholders and is summarised in several published reports [6].

1.4. Understanding change – stakeholder perspectives

Between October 2024 and March 2025, the Change Management Working Group held 20 bilateral meetings with diverse stakeholders (including major EU industry associations, small and medium-sized enterprises (SMEs), contract research organisations (CROs), consultants, non-governmental organisations (NGOs), academia and regulators) to provide input into the roadmap. These bilateral meetings supplemented the above-mentioned stakeholder consultations and the consultations with national representatives at meetings of regulatory committees and expert groups dealing with the legislative areas within the scope of the roadmap. The results of the bilateral meetings are described in Berggren & Worth (2026)[7].

[5] https://webgate.ec.europa.eu/circabc-ewpp/d/d/workspace/SpacesStore/a96d57cb-85a5-48fb-bb03-c751b8e39843/download.

[6] https://single-market-economy.ec.europa.eu/sectors/chemicals/reach/roadmap-towards-phasing-out-animal-testing_en.

[7] Berggren, E. and Worth, A .: Managing change while phasing out animal testing in chemical safety assessments – a collection of stakeholder experiences and expectations, NAM Journal, 2, 2026, p. 100079, https://data.europa.eu/doi/10.1016/j.namjnl.2026.100079,

2. Identifying opportunities towards phasing out animal use

2.1. Recommendation of actions for human health and environmental assessments

A core goal of the roadmap is to provide concrete recommendations for the transition to non-animal approaches across all chemical safety assessment areas. The recommendations reflect the state of knowledge at the stage of drafting the Commission Communication and its accompanying Staff Working Document. Recommendations will need to be adapted as scientific knowledge advances on how to phase out animal testing. They are operationally categorized based on different indicative time horizons for implementation:

1. **Short-term actions** – available approaches for which the process of implementation into regulatory practice, including by omitting redundant tests, can be launched immediately or within a short timeframe.
2. **Medium-term actions** – approaches for which further steps for validation or regulatory adaptation are required before wider use. After finalising required steps, approaches will be implemented immediately when deemed acceptable for use in a given regulatory context.
3. **Long-term actions** – redefining safety assessment paradigms and performance criteria for future non-animal approaches. Long-term actions will include the need to develop further approaches, which will be implemented in a given regulatory context when deemed acceptable. Work on a new scientific safety assessment framework will start immediately as a core activity for implementing the roadmap.

During the development of the roadmap, the working groups of the Commission services with input from stakeholders, created different chapters of the SWD. These chapters contain recommendations for different areas, such as:

1. A long-term change towards a new scientific assessment framework that allows to switch to animal-free regulatory assessments (chapter 2.2.)
2. Overarching opportunities across human health and environmental safety assessments (chapter 2.3.)
3. Environmental safety assessments (chapter 3)
4. Human health safety assessments (chapter 4)

The long-term goal requires development of a new scientific assessment framework, as outlined in the corresponding Commission Communication on the roadmap. Five sub-chapters of the SWD explain the development, conditions and parameters of a new scientific assessment framework further. The first sub-chapter provides background information, gives information on ongoing projects developing a Next Generation Risk Assessments (NGRA), and suggests steps necessary to reach the goal. The second sub-chapter reflects upon reaching the protection goals with NGRAs and the level of confidence needed. The third sub-chapter outlines the path to a hazard classification for human systemic toxicity based solely on the use of alternative approaches to animal testing, while the fourth sub-chapter addresses the switch to a new regulatory system for environmental safety assessments. Finally, sub-chapter five offers reflections about the integration of human health and environmental safety assessments, with respective recommendations for implementation.

The working groups also identified a number of areas that represent overarching opportunities applicable to both human health and environmental safety assessments, which are the following:

1. Replacing *in vivo* with *in silico* methods
2. Developing TK and ADME models for regulatory needs

3. Determining the Point of Departure (PoD) and signatures of toxicity for chemicals through measured omics data
4. Fostering the use of alternatives to animal approaches for nanomaterials and nanoparticles Risk Assessment

Short- to long-term opportunities for replacing, reducing or refining animal testing for environmental safety assessments is described in chapter 3 for the following different areas:

1. Fish acute toxicity
2. Aquatic and terrestrial bioaccumulation
3. Fish chronic toxicity
4. Endocrine disruption
5. Birds and mammals

Finally, chapter 4 provides more detailed considerations on short- to long-term recommendations for moving towards phasing out animal testing for human health safety assessments in the following areas:

1. Acute toxicity
2. Genotoxicity
3. Carcinogenicity
4. Repeated dose toxicity
5. Developmental neurotoxicity
6. Developmental and reproductive toxicity
7. Endocrine disruption – human health

The following tables summarise several actions proposed for (eco)toxicological assessments. The information included in the tables is abbreviated for the purposes of providing an overview and should be read in conjunction with the more detailed chapters of this SWD and the materials referenced therein to understand the recommendations in context. The actions should be taken as illustrative examples.

Table 1: Actions for implementation of alternatives and further development relevant to human health safety assessment

High level goal	Means to achieve it	Specific actions linked	Sector	Short, mid or long term
Replacement	Replacing in vivo study with computational methods	Use of computational models for Acute oral toxicity, Pharmacokinetic modelling	Technical amendments to Annexes VII to X of Regulation (EC) No 1907/2006	Short-term
	"	Use of PK modelling to predict residue depletion in target animal safety studies	Pharmaceuticals (V)	Short-term / mid-term
	Replacing in vivo study with in vitro assays	Use of in vitro assays to predict/measure developmental neurotoxicity, malformations and embryofoetal lethality, pyrogenicity, sensitisation	Cross-sectoral	Short-term / mid-term
Reduction	Waiving of in vivo tests based on historical experience	Use existing information (e.g. on pesticide active substances) to support waiving or deletion of long-term systemic toxicity in 2nd species	Cross-sectoral	Short-term / mid-term
	Reduction	Proposal to reduce and replace animal studies for genotoxicity or carcinogenicity	Cross-sectoral	Mid-term
	Reduction/waiving of studies based on target patient population	Reduction of repeated-dose toxicology (RDT) studies for advanced cancers or severely debilitating/life threatening diseases	Pharmaceuticals (H)	Short-term / mid-term
	Reduction based on additional data *in vitro/in silico* approaches	Reduction through use of complex in vitro models to predict drug induced liver injury / pharmacokinetic parameters / cardiotoxicity / immunotoxicity.	Cross-sectoral (H)	Short-term / mid-term
	Reduction based on *in silico* approaches	Reduction of control animals included in RDT testing through use of virtual control groups	Pharmaceuticals (H)	Short-term / mid-term
	Reduction through study design optimisation	Several approaches (a priori statistical considerations, reductions of dose groups or recovery animals where possible)	Cross-sectoral	Short-term / mid-term

	Reduction through incorporation of multiple readouts in one study	Reduction of *in vivo* studies through inclusion of additional endpoints in repeated dose toxicity studies, based on retrospective data/experience Waiving of long-term *in vivo* studies through shorter omics-enhanced studies	Cross-sectoral	Short-term
Refinement	Refinement of *in vivo* studies where not yet replaceable	Use of evident toxicity rather than lethality as endpoint in acute toxicity studies	Cross-sectoral	Short-term
Establishing an overarching, non-animal scientific assessment framework	Development of new assessment framework	Characterising protection and confidence levels of traditional assessment and non-animal-based scientific assessments based on a new scientific assessment framework	Cross-sectoral	Long-term
	"	Developing a non-animal-based classification system	Cross-sectoral (excluding pharmaceuticals)	Long-term
	"	Describing toxicity through changes measured at molecular level rather than adverse effects in organisms	Cross-sectoral	Long-term
	Frameworks for specific endpoints	To design (for selected endpoints below) a non-animal approach *in vitro/in silico* battery able to reliably distinguish between non-toxic and potentially toxic substances. To obtain qualitative system with high sensitivity suitable for tiered hazard assessment and agree on characterising the endpoint, based on non-animal information feeding a WoE assessment. - genotoxicity - carcinogenicity - reproductive toxicity - endocrine disruption - nanomaterial assessment	Cross-sectoral	Mid-term

Table 2: Actions − further development and implementation of alternatives relevant to environmental safety assessment; the recommendations usually consist of different sub-actions, which can be found in more detail in the respective chapter on recommendations for the environmental safety assessment

High level goal	Means to achieve it	Specific actions linked (described in subsequent SWD chapters)	Short, mid or long term
Acute aquatic toxicity			
Reduction / Replacement	Adapt legislation and guidance	Reduction or replacement based on available methods (*in silico*, *in vitro*, information from other taxa than fish or more sensitive endpoints)	Short-term
Reduction	Approach development	Explore waiving options based on scientific considerations while maintaining an equivalent level of protection	Short-to mid-term
Reduction / Replacement	Frameworks for specific endpoints	Develop an approach for the assessment of acute aquatic toxicity based fully on non-animal approaches	Mid-/ long-term
Bioaccumulation			
Reduction / Replacement	Adapt guidance and legislation	Reduction or replacement based on available methods (*in silico, in vitro*, information from other taxa than fish)	Short-term
Reduction	Approach development	Explore waiving options based on scientific considerations while maintaining an equivalent level of protection	Short-to mid-term
Replacement	Frameworks for specific endpoints	Develop an approach for the assessment of bioaccumulation based fully on non-animal approaches	Mid-to long-term
Chronic fish toxicity			
Reduction	Adapt test guidelines, guidance, legislation	Clarify regulatory requirements for fish testing and waiving options (e.g. using information from other taxa than fish) based on scientific considerations while maintaining the level of protection	Short-term
Reduction / Replacement	Approach development	Gain a better understanding of how to make maximal use of information from existing *in vivo* and non-animal approaches • Link different lines of evidence from non-animal approaches • Develop a reference dataset of *in vivo* chronic fish effects / endpoints inter alia for the generation of Acute to Chronic Ratios	Mid-term

		• Develop a reference dataset relating chronic invertebrate effects to *in vivo* chronic fish effects	
Replacement	Frameworks for specific endpoints	Define a non-animal approach for assessing chronic fish toxicity based on *in silico* and *in vitro* methods	Long-term
Endocrine Disruption (ED) for environmental safety assessments (ESA)			
Reduction / Replacement /Refinement	Adapt test guidelines, guidance, legislation	Adapt test guidelines / guidance / legislation to maximise information from in vivo testing and use tests both for concluding on the ED properties and risk assessment	Mid-term
Reduction	Validation, case studies	Validation of certain *in vivo* methods (EAMA, LATT, FET with thyroid endpoints) as as a measure to potentially reduce animal testing	Mid-term
Reduction / Replacement	Guidance development	Improve guidance to use the OECD conceptual framework rather as toolbox instead of a tiered approach and foster the application of non-animal approaches	Mid-term
Reduction	Research, case studies, guidance development	Link AOP pathways or networks to non-animal approaches; develop guidance on the use of AOP pathways or networks	Mid-term
Refinement	Research, case studies	Map mechanisms and applicability domains of eleutheroembryo methods (RADAR, REACTIV, XETA)	Mid-term
Reduction / Replacement /Refinement	Research, case studies	Gain a better understanding of how to make maximal use of information from existing non-animal approaches, including those for the assessment of human health, and by linking different information from non-animal approaches (different sub-actions, see chapter on ED for ESA)	Mid-term / long-term
Replacement	Research, case studies	Build evidence on cross-species extrapolation by exploring key events/AOPs for different species	Mid-term / long-term
Replacement	Frameworks for specific endpoints	Define a framework for assessing endocrine disruption for environmental safety with a mechanism-based panel of *in silico* and *in vitro* assays taking into account knowledge of (quantitative) Adverse Outcome Pathways	Long-term
Long-term change for environmental safety assessment for all taxa (including mammals and birds)			
Replacement	Development of new assessment framework	Develop animal-free environmental Next Generation Risk Assessment (NGRA) and implement it for environmental safety assessments.	Long-term

Harmonise human health and environmental safety assessment under One Health		It is important to explore the links between environmental safety assessments and human health NGRA approaches under a one-health concept to ensure the best use of data.	Mid- / Long-term

2.2. Preparing for the transition to animal-free chemical safety assessment

2.2.1. The long-term shift towards "Next-Generation Risk Assessment" (NGRA) in EU chemicals legislation

Terminology

At multiple occasions during the development of the roadmap, e.g. at the workshops organised by the Commission, but also in the working groups for human health or environmental safety assessment, the terminology questions were raised. This includes definitions for test methods without the use of animals or for "animal" itself, the term "Next-Generation Risk Assessment" (NGRA)". While definitions are important to clarify the subject of a discussion, these definitions often contain context-specific elements, and a universally accepted "one-size-fits-all" definition is not available. Without claiming to be exhaustive, terminology used to signify a test system that is not using animals or that is limited to species not considered sentient includes

- Alternative test methods (to animal testing), alternative approaches or alternative test systems
- Non-animal methods, non-animal testing (system), non-animal approaches, or non-animal data
- Animal-free methods or animal-free approaches
- New Approach Methodologies (NAMs)

EU legislation, while not defining the terms, is using the terms mentioned under the first two bullets above[8], while scientific literature is commonly using the term NAM, albeit with different meanings.

Throughout the roadmap Commission Communication and the Staff Working Document, preference is given to:

- Non-animal approach, signifying methods or approaches that do not require the use of any live animal including in their early life stages. They comprise *in vitro* and *in silico* methods (including read-across and grouping), organs-on-a-chip etc.
- Alternative method or approach to animal testing, signifying every method or approach that supports replacing, reducing or refining animal testing (Three Rs). As such, it can involve animal testing if it leads to a reduction or refinement of animal testing.

Directive 2010/63/EU contains a Definition of the Three Rs:

- **Replacement** – a method or testing strategy not entailing the use of live animals.
- **Reduction** – reducing the number of animals used in a project for a certain purpose without compromising the project's objectives
- **Refinement** – refinement of methods to eliminate or reduce pain, suffering, distress or harm to animals. The avoidance of testing of sentient species[9], using instead species with a potentially lower capacity to experience pain, suffering and distress, or of early life stages of vertebrates or amphibians would therefore result in refinement.

It is important to recognise that both reduction and refinement remain important stepping stones towards the goal of phasing out animal testing. By its very nature, however, it is also clear that **the roadmap aims at ultimately relying on non-animal approaches only.**

[8] E.g. the Regulation on Plant Protection Products (EC) No 1107/2009, Directive 2010/62/EU on the protection of animals used for scientific purposes, or the REACH Regulation (Regulation (EC) 1907/2006)

[9] It is acknowledged that the question which species can be considered "non-sentient" e.g. fruit flies, water fleas, round worms and embryos of zebrafish and frogs as explored by the ASPIS cluster project PrecisionTox is the subject of an ongoing scientific debate beyond the scope of this text. The term is nevertheless used here to clarify that new *in vivo* testing in species that have been identified as "sentient" should not be performed under NGRA.

Last not least, EU legislation does not define the term "animal". Directive 2010/63/EU applies to the animals and life-stages listed in its Article 1, which excludes for instance animals in their early life stages (eleuthero embryos) or invertebrates. Nevertheless, animals in their early life stages or invertebrates are naturally considered "animals". Consequently, the term "non-animal approach" excludes testing with animals in their early life stages, invertebrates etc.

The roadmap therefore assumes that in the long-term also testing with animals in their early life stages or invertebrates need to be replaced to reach the ultimate goal of phasing out animal testing for chemical safety assessments. However, in the short- to medium-term, testing with animals (vertebrates) in their early life stages or with invertebrates is considered an refinement and to be acceptable, if it reduces the reliance on vertebrate testing beyond early life stages. This consideration takes into account that suffering of animals differs.

Next-Generation Risk Assessment based on alternatives to animal testing

A similarly recurring issue in the discussions around the innovation of chemical risk assessment[10] is the request for clear definitions of the term "Next-Generation Risk Assessment" (NGRA). Various authors have made attempts at providing direct or indirect definitions for the term, e.g. Dent et al. (2018), Berggren and Worth (2023), Marx-Stoelting et al. (2023), Sewell et al. (2024) or Herzler et al. (2025), to name just a few. While the desire for exact definitions is understandable (and such definitions may eventually be formulated), they are not a prerequisite[11] for constructing and implementing the roadmap.

The following considerations have been discussed in the context of NGRA for the assessment of human health. Similar considerations on NGRA for the environmental safety assessment are presented in chapter. 2.2.4. Long-term change - A new paradigm for regulatory environmental safety assessments.

As presented by members of the Commission working group on Human Health at the 3rd Roadmap Conference in June 2025, the eventual full phasing out of animal testing for chemical safety assessments requires the transition to a "New Safety Paradigm" characterised by the following elements:

- human-centric safety assessments[12] using alternatives to animal testing;
- mechanistic understanding to distinguish relevant from irrelevant substance effects;
- utilisation of up-to-date science;
- high throughput and efficiency;
- a high level of protection and confidence, demonstrated explicitly and transparently;
- consistency across regulatory sectors in line with the "One Substance, One Assessment" (OSOA) principle;
- integration with environmental safety in line with the "One Health" paradigm.

[10] Note that "chemical risk assessment" is used here as a collective term inclusive of partial aspects of hazard or exposure assessment and does not only refer to "full-scale" risk assessments.
[11] Of note, even pivotal NGRA-related reports by US EPA did not attempt to provide such a definition: (2014). *Next generation risk assessment: incorporation of recent advances In molecular, computational, and systems biology (final report)* (EPA/600/R-14/004). https://ordspub.epa.gov/ords/eims/eimscomm.getfile?p_download_id=520491 and Thomas, R. S., Bahadori, T., Buckley, T. J., Cowden, J., Deisenroth, C., Dionisio, K. L., Frithsen, J. B., Grulke, C. M., Gwinn, M. R., Harrill, J. A., Higuchi, M., Houck, K. A., Hughes, M. F., Hunter, E. S., III, Isaacs, K. K., Judson, R. S., Knudsen, T. B., Lambert, J. C., Linnenbrink, M.,Williams, A. J. (2019). The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. *Toxicological Sciences*, *169*(2), 317-332. https://doi.org/10.1093/toxsci/kfz058 .
[12] It should be noted that efficient testing strategies that provide information for both human health and environmental assessments might need to make compromises as regards the use of more human centric approaches. Elements for a NGRA for environmental assessments are discussed in Chapter 2.2.4.

For the purpose of this document, "NGRA" then refers to chemical risk assessment conceptual frameworks or practical workflows implementing one or more elements of this paradigm. In an interim phase, NGRA might also rely on animal testing that supports reducing or refining animal testing, if that testing addresses other aspects of the new paradigm (as exemplified by the augmentation of "classical" *in vivo* rodent studies augmented with transcriptomics measurements to elucidate possible mechanisms/modes of action). However, the long-term goal remains to use non-animal approaches only.

The novelty of this paradigm can best be understood in contrast to present-generation chemical risk assessment frameworks/workflows which are largely characterised by

- animal-centric testing and assessment strategies;
- empirical "black box" testing in whole organisms mostly without mechanistic understanding at the molecular, cellular, organ or system level;
- slow uptake of new scientific developments;
- slow and resource-intensive testing, often following an untargeted "shotgun" approach;
- an implicit and therefore often untransparent level of protection, with uncertainties not considered (or considered, but not reported) in a quantitative way;
- variability of concepts across regulatory sectors;
- compartmentalisation between human health and environment.

Historical development and future perspectives

United States, Canada and the UK

More than a decade ago, the United States Environment Protection Agency (US EPA) published their pivotal report "Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology" (US EPA, 2014), and initiated the Tox21 programme, a joint effort of several U.S. federal agencies. This programme provided the backbone for introducing a paradigm change in science-based chemical risk assessment in the U.S., with worldwide repercussions. A more detailed vision and strategy was provided in the landmark report "Toxicity Testing in the 21st Century. A Vision and a Strategy", published by the U.S. National Academy of Sciences (National Academies, 2007), and followed by the formal installation of an alliance of the involved U.S. agencies in 2008. This co-operation has provided an unprecedented dynamic for the science-based innovation of chemical risk assessments.

In 2018, ICCVAM, the United States Interagency Coordinating Committee on the Validation of Alternative Methods, published its "Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States" (ICCVAM, 2018). Most recently, in January 2024, the US National Institutes of Health's "Complement-ARIE" fund (Sunderic et al., 2025) was approved, which will provide a budget of $35-40 M per year over the next ten years, *inter alia* for technology development projects/centres, a data and resource coordinating centre, a validation network for regulatory implementation, community engagement and training and strategic engagement.

In April 2025, the US Federal Drug Administration (US FDA) released its "plan to phase out animal testing requirement for monoclonal antibodies and other drugs".[13] In June 2025, the US National Institutes of Health followed with the announcement to reduce use of animals in NIH-funded research

[13] https://www.fda.gov/news-events/press-announcements/fda-announces-plan-phase-out-animal-testing-requirement-monoclonal-antibodies-and-other-drugs, last accessed 2025-07-12

by focusing its funding programs on "human-based research technologies".[14] In July 2025, Canada published its Strategy to replace, reduce or refine vertebrate animal testing under the Canadian Environmental Protection Act[15] and in November 2025, the UK published its Replacing animals in science strategy[16].

European research projects and partnerships

In Europe, major EU projects, such as Predict-IV (Mueller et al., 2015), SEURAT-1 (Daston et al., 2015; Gocht et al., 2015) or EU-ToxRisk (Daneshian et al., 2016; Moné et al., 2020) have achieved great progress on the scientific side and have successfully started addressing (and bridging) the communication gap between academic researchers and regulatory practitioners. Several EU projects active at the time of drafting this document, such as the ASPIS[17] cluster featuring the human-health-related projects RISK-HUNT*3R*[18], ONTOX[19] and PrecisionTox[20], the ENKORE[21] cluster with its projects EDC-MASLD, MERLON, HYPIEND and ENDO-MIX, or the EURION[22] cluster with its projects ATHENA, EDCMET, ENDpoiNTS, ERGO, FREIA, GOLIATH, OBERON and SCREENED, focusing on the assessment of endocrine disruptors, are building on and complementing this work.

i) ASPIS cluster

A core initiative within the ASPIS cluster is the ASPIS Alternative Safety Profiling Approach (ASPA), a structured pragmatic workflow for NGRA developed within the RISK-HUNT*3R* project which provides a tiered decision-making framework (see below for more details).

The ONTOX project develops physiological maps, quantitative adverse outcome pathway networks and ontology frameworks with the aim of providing artificial intelligence (AI)-assisted probabilistic NGRA workflows.

The PrecisionTox project, based on the hypothesis that fundamental mechanisms of toxicity have been conserved between species, *inter alia* applies omics technologies to non-traditional model organisms (e.g. invertebrates) to predict human-relevant toxicological effects. This cross-species approach seeks to bridge gaps between environmental and human risk assessment, reinforcing the "One Health" concept.

ii) European Partnership for the Assessment of Risks from Chemicals (PARC)[23]

The Partnership for the Assessment of Risks from Chemicals (PARC) contributes to the development of NGRA strategies on diverse levels, from Adverse Outcome Pathways (AOPs) to strategies for legal implementation (Herzler et al., 2025). The NGRAroute activity, a strategic initiative aimed at establishing animal-free NGRA as the default framework under EU chemicals legislation, seeks to integrate NAMs into regulatory workflows while addressing scientific and policy challenges. Key priorities cover fundamental conceptual questions around integrating diverse NAMs into workflows that are still able to provide classical regulatory outputs such as classification and labelling, points of

[14] https://www.nih.gov/news-events/news-releases/nih-prioritize-human-based-research-technologies last accessed 2005-07-12
15 Strategy to replace, reduce or refine vertebrate animal testing under the Canadian Environmental Protection Act, 1999 (CEPA) - Canada.ca
16 Replacing animals in science strategy - GOV.UK
[17] https://aspis-cluster.eu
[18] https://www.risk-hunt3r.eu
[19] https://ontox-project.eu
[20] https://precisiontox.org
[21] https://enkore-cluster.eu/
[22] https://eurion-cluster.eu/
[23] https://eu-parc.eu

departure, reference/guidance values etc. In addition, the focus of the activity is on the legal uptake and implementation side, with its position at the border between science and policy.

Based on the scientific progress in recent years, the understanding of possible ways to build an integrated, overarching NGRA framework has greatly increased over the past decades. NGRAroute has tried to summarise these experiences and to bundle them into ten guiding principles for a future NGRA framework, which were first discussed during the 1st Roadmap workshop in December 2023. Following an intensive stakeholder commenting round, a revised version was later provided as an Additional Deliverable[24] from the PARC Partnership:

1. **Protection**: The framework allows to determine whether an adequate level of protection for human health and the environment is attained.
2. **Confidence**: The framework allows to determine whether an adequate[25] level of confidence is attained, especially when concluding on the absence of relevant hazard, exposure or risk.
3. **Generation of information**: The framework takes into account all relevant existing information. If additional information is required, it is generated without *in vivo* testing in sentient species.
4. **Science**: The framework uses adequate and reliable modelling, testing and assessment methodology, with high scientific relevance to the protection of human health and the environment, representing the current state of scientific knowledge.
5. **Evidence integration**: The framework can integrate all relevant lines of evidence with acceptable reproducibility.
6. **Regulatory workflows**: The framework is applicable to all relevant chemical hazard, exposure and risk assessment workflows.
7. **Biology**: The framework covers all relevant (eco)toxicological pathways and endpoints.
8. **Exposure**: The framework covers all relevant exposure levels, durations and routes.
9. **Chemistry**: The framework is applicable to substances in their relevant physical forms, their transformation products, groups of substances, mixtures[26] and to substances in articles.
10. **Efficiency**: The framework allows for assessments within an acceptable time- and cost-frame. It includes integrated testing and assessment approaches that are as complex as scientifically necessary, but also as simple and straightforward as possible.

The 10 guiding principles were proposed as a means to:

- foster broad consensus on fundamental questions at an early developmental stage instead of getting into lengthy discussions on small details early on;
- define the political, scientific and regulatory boundaries of NGRA; thereby helping to
- structure the further roadmap work and provide a basis for focused topical discussions.

It is noted that the principles describe an ideal chemical risk assessment framework, with some of them also not well addressed by present-day risk assessment schemes. In this regard, the principles may also be used to highlight possible gains to be made by switching from the current to an NGRA-based regulatory system.

Following the publication of the roadmap, it is planned to further develop the NGRAroute activity into a platform for NGRA implementation in the EU. In this context, the project also intensified work on Next-Generation Environmental Risk Assessment.

[24] https://www.parcopedia.eu/wp-content/uploads/2024/10/PARC_AD2.1_submitted_approval_pending.pdf
[25] as defined by the respective regulatory context and problem formulation
[26] as defined by Art. 3 of the REACH Regulation

Broad consensus among all stakeholders will be key for acceptance of the shift towards NGRA. With the intention to create a hub bringing together the various factions of the chemical risk assessment community for an open dialogue aiming at broad consensus solutions, PARC has created the online knowledge management and community platform PARCopedia.[27]

iii) EPAA NAM Designathon

The EPAA NAM Designathon is a collaborative initiative aimed at developing a novel framework for classifying for human systemic toxicity using NAMs (or alternatives to animal testing). The project seeks to transition from traditional animal-based classification criteria to non-animal systems while maintaining or enhancing protection levels.

The Designathon focuses on creating a systemic toxicity classification matrix based on alternatives to animal testing, designed to replace animal-derived criteria with mechanistic and kinetic data, targeting a broad basis of human systemic toxicity endpoints (e.g. CMRs, endocrine disruption, STOT), but without focussing necessarily on the current classification. By calibrating the new system based on already assessed chemicals, the initiative aims to ensure equivalent level of protection (i.e. the same risk management decisions can be made) and to inform future UN GHS and broader EU chemical legislation revisions.

The framework employs a classification matrix based on three levels of concern (low, medium, high concern) using only toxicokinetic (TK) and toxicodynamic (TD) properties. It prioritises flexibility, allowing integration of future regulatory needs and emerging alternative approaches. Though initially focused on human health, the framework is designed for extension to environmental safety.

EU regulatory bodies

In addition to the large-scale research projects named above (and others), regulatory bodies and agencies have started initiatives to actively investigate and promote alternatives to animal testing and NGRA based on alternative approaches.

i) Scientific Committee on Consumer Safety (SCCS)

In the EU, the Scientific Committee on Consumer Safety (SCCS), driven by the ban on animal use for the testing of cosmetic ingredients, has frequently adapted its Notes of Guidance with a view to non-animal approaches and NGRA (most recently in SCCS (2023)).

The EU Cosmetic Products Regulation (No 1223/2009) mandates that all cosmetic products must be safe, with safety evaluations based exclusively on data generated without using animal testing. This prohibition on animal testing extends to marketing bans, making the adoption of non-animal approaches unavoidable. The SCCS framework prioritises exposure-driven risk assessment, integrating hazard data, chemical characterisation, and exposure estimates. Key elements include:

- Validated non-animal methods: Priority is given to OECD-validated methods, though scientifically valid non-validated non-animal approaches (criteria under development) are also considered;
- Notes of Guidance: The SCCS's 12th revision outlines testing requirements, with forthcoming updates addressing NGRA guidance;
- Threshold of toxicological concern (TTC): a risk assessment tool for low-exposure scenarios.

The NGRA approach integrates various data sources, including *in silico*, *in chemico*, and *in vitro* studies, along with read-across and historical *in vivo* data. This strategy is designed to address gaps in systemic toxicity assessments, particularly in areas lacking (validated) tests for repeated-dose toxicity. A "safe harbour" approach allows for the parallel submission of NGRA alongside standard regulatory dossiers.

[27] https://parcopedia.eu

This dual-submission framework aims at fostering regulatory acceptance of novel methodologies while ensuring compliance with existing standards.

A weight of evidence (WoE) approach strengthens confidence in non-animal approaches by integrating multiple data streams. This helps mitigate the limitations of individual assays. WoE is considered particularly robust when it aligns with established AOPs and key biological events, providing a structured, scientifically sound basis for regulatory decision-making.

In terms of the development of non-animal-based NGRA, the SCCS advocates for:

- harmonised criteria for physiologically based pharmacokinetic (PBPK) modelling in NGRA.
- validation or scientifically valid status of non-animal approaches used in existing NGRA case studies, many of which derive from contract research organisations (CROs) but lack formal regulatory recognition.
- Differentiation between non-animal approaches and broader new approach definitions (in the sense of alternatives to animal testing) to avoid ambiguity.

The SCCS framework balances regulatory rigour with scientific innovation, ensuring cosmetic safety through quantitative risk assessment rather than hazard alone. By promoting NGRA, WoE, and exposure-driven strategies, the Committee aims to advance animal-free methodologies while maintaining consumer trust. Continued collaboration across sectors and refinement of validation criteria will be pivotal to overcoming current limitations and achieving harmonised implementation.

Global developments

i) APCRA

As an important example of global collaboration among regulators from US, Canada, Europe, Asia and Australia, the initiative for "Accelerating the Pace of Chemical Risk Assessment[28]" (APCRA) aims to discuss *"progress and barriers in applying new tools to prioritization, screening, and quantitative risk assessment of differing levels of complexity"* and *"opportunities to increase collaboration in order to accelerate the pace of chemical risk assessments"* by means of case studies with high regulatory relevance (Kavlock et al., 2018; Paul Friedman et al., 2019; Paul Friedman et al., 2025).

ii) OECD

On a global scale, the Organisation for Economic Co-Operation and Development (OECD), supported by its member countries, with large contributions from EU Member States, the European Commission's Directorate-General Joint Research Centre (JRC) including the EU Reference Laboratory for alternatives to animal testing (EURL ECVAM), and European agencies like the European Chemicals Agency (ECHA) and the European Food Safety Authority (EFSA), has successfully facilitated the conceptual development of innovative chemical hazard and risk assessment methodologies, such as the Defined Approaches for Skin Sensitisation (DASS) Guideline Document 497 published in 2021 (OECD, 2023a) or the Recommendations on Evaluation of Data from the Developmental Neurotoxicity (DNT) In-Vitro Testing Battery (OECD 2023c). Via the "Integrated Approaches to Testing and Assessment" (IATA) case studies project, the OECD continues to actively explore their practical application in the regulatory context. The OECD also plays a key role in further developing concepts, e.g. the concept of Adverse Outcome Pathways (AOPs) and practical approaches to validate test as well as non-test methods regarding their fitness-for-purpose for regulatory chemical risk assessment and by developing relevant guidance in the field (e.g. (OECD, 2017a, 2017b, 2017c, 2023b).

[28] Home - APCRA

Towards an integrated, overarching NGRA framework

The need for an overarching framework

Current chemical risk assessments frameworks feature a number of traditional concepts, e.g. substance-by-substance assessments or deterministic hazard/exposure/risk characterisation, as well as traditional test methods using experimental animals that have been challenged for, among others, ethical, scientific and performance-related reasons.

To phase out animal testing from chemical risk assessments, as desired by the roadmap, there can in principle be two strategies:

1. The chemical risk assessments framework as such is maintained, but the (animal-based) test methodology is replaced one by one - with non-animal approaches, where possible.

2. A new ("next-generation") risk assessments framework is introduced.

Regarding 1., ECHA has noted that

"Until recently, NAMs development aimed to fully replace animal testing for each specific regulatory endpoint. These developments have been successful for some relatively simple endpoints (like skin sensitisation), where the adverse effect and the mechanism(s) leading to this effect are relatively well understood. Development of NAMs for more complex endpoints has so far been less successful.
By now, the scientific community and regulators widely accept that it would be almost impossible to develop one-to-one replacements of animal tests by NAMs for more complex endpoints such as e.g., repeated dose toxicity or reproductive/ developmental toxicity." (ECHA, 2025)

As a consequence, and in line with 2. above, it has become increasingly clear that not only the methods, but also the risk assessment frameworks themselves need to be adapted to accommodate new paradigms and concepts and to respond to regulatory challenges not well addressed by the current frameworks, such as the efficient assessment and management of a huge number of chemicals on the market, possible combination effects, new types of chemicals and materials or chemical effects on biodiversity.

In addition, the development of a new, unified NGRA framework also offers the chance to better integrate human health and environmental risk assessment workflows across the diverse chemical sectors, resulting in a more holistic approach in the spirit of the "One Substance, One Assessment" philosophy formulated *inter alia* in the EU Chemicals Strategy for Sustainability (European Commission, 2020). Changing the current chemical risk assessment framework, however, comes at a price, as it will trigger the need for deeper changes in the current chemicals' legislation and the accompanying guideline and guidance documents, compared to a one-by-one test method replacement strategy.

The necessary first step in this direction consists of elaborating the conceptual details of a new, integrated, overarching NGRA framework, to better understand where (and which) legislative and other fundamental conceptual changes are needed. In addition, by structuring the framework and associated workflows, current knowledge gaps, where applicable, can be identified and targeted by future funding of corresponding research initiatives.

Chemical risk assessment workflows and outputs

A large variety of overarching as well as sector-specific EU chemicals legislation is in place and covers a wide range of different chemicals, uses and exposures (see Annex I). While chemical risk assessment under some of the sectorial regulations may sometimes address issues specific for that sector (e.g. certain unique exposure scenarios), in general only a limited set of basic risk assessment workflows will be used across all sectors in the vast majority of cases, such as:

- determination of the fate of a chemical in the human body (absorption, distribution, metabolism, excretion, bioaccumulation) or the environment (biodegradation partitioning, ADME, etc.);

- hazard identification/classification and labelling;
- hazard characterisation, i.e. identification of points of departure (PoDs) for risk assessment, derivation of safe exposure levels (health-based or environmental guidance values environmental quality standards, MRLs), including extrapolation across species, duration, and routes;
- exposure assessment (estimation of migration from products, single-source or aggregate internal human exposure in serum/plasma or in target tissues, in environmental compartments etc.);
- screening/prioritisation of substance of high concern;
- full-scale risk characterisation / exposure-driven safety assessment.

Under the current chemical risk assessment system, each of the above workflows is designed to provide specific, "tangible" outputs needed to address the risk assessment problem at hand. In addition, these workflows build on one another. Therefore, a future unified NGRA framework should in the end be able to integrate all relevant human health and environmental safety assessment workflows into a modular scheme in which some of the modules may or may not apply, depending on the specific chemical risk assessment problem formulation at hand.

Understanding the outputs required by chemical risk assessment, e.g. an estimate of internal exposure, a safe dose or a measure of risk for humans and the environment, and how these outputs can be delivered, is an essential prerequisite for creating the future NGRA framework based on non-animal approaches.

Standardised NGRA workflows: the ASPIS Alternative Safety Profiling Approach (ASPA)

One of the overall key success factors of the current chemical risk assessment frameworks lies in the highly schematic and standardised nature of the regulatory workflows built around the regulatory outputs described above. This standardisation as implemented by test guidelines and guidance at the sub-legal level, is crucial for:

- providing a common understanding between actors with multi-disciplinary backgrounds;
- the global harmonisation of chemical risk assessment;
- schematic training including an efficient transfer – via guidance – of knowledge from specialised experts to risk assessors faced with the need to cover a broad range of interdisciplinary scientific questions;
- comparability of assessments and creating level playing fields for all market actors.

The ASPIS cluster, and specifically the ASPIS cluster project RISK-HUNT*3R*, is currently developing the so-called "Alternative Safety Profiling Approach", or ASPA. The ASPA attempts to lay out a modular NGRA framework as described in the previous section for addressing the outputs required by hazard/exposure/risk assessment, using only non-animal approaches. More specifically, it aims to standardise the processes for data collection, data integration and data reporting. Operationalisation of NGRA, i.e. the use of non-animal approaches for regulatory purposes, requires that outcome(s) are provided with sufficient certainty, that it is fully clear which non-animal approaches and tools were used and that all intermediate and final decision points within a tiered testing strategy process are transparently justified and documented. In other words, full implementation of NGRA – and thus phasing out animal testing for chemical risk assessment – would be greatly accelerated when these aspects are addressed in a transparent and structured way.

ASPA is a well-guided, adaptable workflow co-developed by a large group of scientists, and improved stepwise by stakeholder input and application in defined case studies. A first overview of the ASPA workflow, including its principles, structural outline, and its operability is given in Leist et al. (2025). Depending on the specific problem formulation, only a subset of the ASPA elements may apply. Accompanied by a dedicated software tool, ASPA-*assist*, one of the main aims of the ASPA is to guide

users through an NGRA workflow specific for their respective problem formulation and to support them in deciding whether, and if so, which further information (derived from non-animal approaches) is needed to come to a conclusion. Guidance is given by different information layers within each step of ASPA. Also, the ASPA-*assist* provides a detailed report of the assessment done.

Key components of NGRA in ASPA include:

- **Hypothesis-driven approach**: Data is collected based on a predefined hypothesis, ensuring relevance and efficiency;
- **Exposure and kinetics:** Understanding how chemicals interact with biological systems at different exposure levels (this includes ADME properties);
- **Hazard identification and characterisation:** The use of high-throughput methods, including *in silico* (QSAR, structure-based modelling) and *in vitro* (mechanistic assays, cell stress reporters, transcriptomics), to determine potential toxic effects and the dose at which these may occur;
- **Decision logic and risk assessment:** Integrating deterministic and/or probabilistic models to determine the likelihood of hazard and exposure, thereby refining risk predictions.
- **Case studies**: ASPIS is validating ASPA through diverse case studies across chemical sectors, including CLP classification and risk assessments.
- **Probabilistic risk assessment:** By modelling uncertainty ranges for hazard and exposure data, ASPA provides regulators with confidence intervals for risk outcomes, facilitating informed decision-making.
- **Evolutionary toxicology:** PrecisionTox contributes with omics data to identify molecular toxicity pathways shared across animal species , linking toxicogenomic responses in invertebrates to human AOPs, thereby bridging ecological and human health assessments.
- **Stakeholder collaboration:** Workshops engage regulators, academia, and industry to refine workflows and address implementation barriers.

While ASPA is still under construction, it is already clear that it can become of paramount relevance for the implementation of the roadmap, because it may *inter alia* serve:

- to demonstrate how a future animal-free NGRA framework can produce the same or equivalent outputs to those provided by the current, largely animal-based framework;
- to provide a concrete blueprint for NGRA-based strategy to be incorporated into the legislations;
- to help identify, by means of dedicated case studies, knowledge or methodological gaps in the workflow, which can then be addressed in a targeted way by the Commission working groups, and to support these groups in developing concrete decision criteria regarding the acceptability of outputs generated by non-animal approaches and associated uncertainties.

Formulating NGRA frameworks such as the ASPA is an indispensable prerequisite for the further work on the Commission roadmap: only if the details of such a framework have become sufficiently clear can the need for changing the legislation and/or respective guideline and guidance documents be formulated.

Stepwise, tiered strategies proposed in ASPA for assessment and generation of new information under NGRA

A future NGRA framework as the basis for chemical risk assessments based on non-animal approaches will likely need to take into account data from different methods, including relevant (as defined by the problem formulation and regulatory context) existing data, and combine them in a stepwise approach. During the time of transitioning to chemical safety assessments without the use of animals, it may still be necessary to use *in vivo* testing where the problem formulation cannot be addressed adequately by use of existing data or new data generation with non-animal approaches. In the spirit of reducing or

refining animal testing, additional information might be generated with *in vivo* testing with less sentient species, e.g. conducting testing with early life stages.

The following scheme[29] [30], without claiming to be exhaustive, describes potential elements of a stepwise strategy for chemical risk assessments towards phasing out animal testing based on a synopsis of approaches published so far.

Step 1 Assess and integrate all existing and relevant hazard and exposure data of sufficient quality, in order to identify any relevant knowledge gaps as well as any relevant regulatory concerns triggering the need for further data generation. In cases where biological mechanisms or pathways are conserved between humans and other species, it could also be justified – subject to a case-by-case analysis – to use existing data generated for environmental assessment in non-mammalian species to fill data gaps when evaluating human health (and similarly using mammalian, including human, data to support environmental safety assessments).

Already in the near future, machine-learning or other AI methods may dramatically improve the consideration of existing toxicity or exposure data (provided that the required high-quality curated datasets are generated and become publicly available).

Step 2 Follow up on such gaps or triggers first by means of read-across from similar substances, where possible, see e.g. Patlewicz and Shah (2023)[31].

Step 3 Depending on the problem formulation and regulatory context, use filters based on physico-chemical, *in silico* models or ADME [32] properties to rule out relevant exposure, where possible (to be further explored), taking into account all potentially relevant chemical species (parent, metabolites, abiotic degradation products).

Step 4[33] Next, if possible or required, perform a bioactivity screening step using *in chemico*, *in silico* or *in vitro* models to identify possible relevant pathways of biological perturbation. This bioactivity screen is meant as a gatekeeper to focus further testing on any triggers that might have been detected during the previous steps and therefore, it

- needs to have a high biological coverage, to ensure that no relevant bioactivity is missed,
- needs to be "protective", i.e. it must have a high negative predictive value (likelihood that a negative outcome really is a negative), thereby providing high confidence when concluding that no relevant hazard/exposure/risk is present, and
- should preferably also have a high positive predictive value to keep potential positive follow-up testing (see next steps) within reasonable limits.

Step 5 Where possible and appropriate, follow up on any triggers from the screening approach or remaining gaps using non-animal approaches. During the transition period towards an animal-free chemical safety assessment, targeted *in vitro* testing or *in vivo* testing with a preference for testing with non-sentient animals might still be necessary.

[29] It is noted that the approach presented here is written from a human-health perspective. Also, for environmental risk assessments, new ways of tiered information generation e.g. based on monitoring data could be considered

[30] A similar stepwise approach forms the foundation of the ASPA

[31] The EU agencies published guidance on read-across, e.g. Guidance on the use of read-across for chemical safety assessment in food and feed - - 2025 - EFSA Journal - Wiley Online Library

[32] Absorption, metabolism, distribution, excretion

[33] Within this step - in analogy to current risk assessment practice – a tiered approach could be applied, starting with less resource-intensive approaches the results of which have been demonstrated to be on the conservative side and then refining with more complex approaches only if necessary.

The methodology used should be "predictive", i.e. it should have high predictive accuracy (i.e. a high fraction of correctly identified positive and negatives) regarding adverse outcomes for human health (or other species or populations in case of environmental safety assessments). It should allow hazard classification, where required, cover all relevant mechanistical pathways, and consider the need for metabolic competence to account for relevant transformation products.

It should also provide dose/concentration-response information to derive points of departure (PoDs) for further risk assessment, where this is required by the problem formulation and regulatory context.

Step 6 (Only during the transition to NGRA:) Perform new *in vivo* testing in sentient animals after a case-by-case assessment concluding that the remaining knowledge gaps need to be closed because a) there is reason to assume that closing them will have a relevant impact on risk management, that b) alternative methodology is inadequate for addressing the knowledge gap, and c) there is a legal requirement to close the knowledge gap. If all of this is the case, the *in vivo* testing strategy pursued should minimise suffering and stress of the experimental animals to the maximum extent possible. It should also make best use of the animals, e.g. by adding relevant alternative testing components, such as omics (cf. concept of "TG+" studies, PARC (2023)).

In the above concept, it is critical to define after each step whether the knowledge obtained is already sufficient to inform risk management (and therefore the assessment can stop). Non-animal approaches should effectively be used to determine the presence or absence of hazards, ultimately resulting in classification without testing on animals. For this, clear decision criteria are needed, in particular for reliably demonstrating "no relevant toxicity" (to be defined by the problem formulation and regulatory context).

While the knowledge required for the hazard, exposure or risk assessment, as well as the potential areas of concern needs to be defined *a priori*, and a catalogue of established methods and models may be provided, the latter steps of the stepwise approach may also require case-by-case decisions which would need to be standardised as much as possible to avoid subjectivity and bias, unnecessarily lengthy assessments and delayed decisions due to the tiered approach. *Inter alia* PARC conducts projects that can provide insights on this matter.

Incorporating a stepwise approach, such as the one described above, would likely require a revision of legislative texts[34].

During the transition period to NGRA, *in vivo* testing will still be required for some time. In the current risk assessment framework, however, a catalogue of standard *in vivo* test requirements is often in place, while the likelihood that such tests will provide added value to risk management is not clear in an individual case. To better address this aspect already while *in vivo* testing cannot be completely replaced for certain areas of concern, criteria and methodology should be developed to determine the likelihood that additional *in vivo* testing in animals will provide added value for risk management.

[34] This SWD does not intend to pre-empt any decisions or discussions needed that could lead to a legislative proposal.

References

Berggren, E., & Worth, A. P. (2023). Towards a future regulatory framework for chemicals in the European Union – Chemicals 2.0. *Regulatory Toxicology and Pharmacology*, *142*, 105431. https://doi.org/10.1016/j.yrtph.2023.105431

Daneshian, M., Kamp, H., Hengstler, J., Leist, M., & van de Water, B. (2016). Highlight report: Launch of a large integrated European in vitro toxicology project: EU-ToxRisk. *Archives of Toxicology*, *90*(5), 1021-1024. https://doi.org/10.1007/s00204-016-1698-7

Daston, G., Knight, D. J., Schwarz, M., Gocht, T., Thomas, R. S., Mahony, C., & Whelan, M. (2015). SEURAT: Safety Evaluation Ultimately Replacing Animal Testing - recommendations for future research in the field of predictive toxicology [journal article]. *Archives of Toxicology*, *89*(1), 15-23. https://doi.org/10.1007/s00204-014-1421-5

Dent, M., Amaral, R. T., Da Silva, P. A., Ansell, J., Boisleve, F., Hatao, M., Hirose, A., Kasai, Y., Kern, P., Kreiling, R., Milstein, S., Montemayor, B., Oliveira, J., Richarz, A., Taalman, R., Vaillancourt, E., Verma, R., Posada, N. V. O. R. C., Weiss, C., & Kojima, H. (2018). Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. *Computational Toxicology*, *7*, 20-26. https://doi.org/10.1016/j.comtox.2018.06.001

ECHA. (2017a). *Non-animal approaches: Current status of regulatory applicability under the REACH, CLP and Biocidal Products regulations* (ECHA-17/R/24/EN).

ECHA. (2017b). *Read-Across Assessment Framework (RAAF)* (ECHA-17-R-01-EN).

ECHA. (2023). *Report on the European Chemicals Agency's "New Approach Methodologies Workshop: Towards an Animal Free Regulatory System for Industrial Chemicals". 31 May – 1 June 2023, Helsinki, Finland*.

ECHA. (2025). *Key areas of regulatory challenge*. https//doi.org/10.2823/8572710

Escher, S. E., Partosch, F., Konzok, S., Jennings, P., Luijten, M., Kienhuis, A., de Leeuw, V., Reuss, R., Lindemann, K.-M., & Bennekou, S. H. (2022). Development of a Roadmap for Action on New Approach Methodologies in Risk Assessment. *EFSA Supporting Publications*, *19*(6), 7341E. https://doi.org/10.2903/sp.efsa.2022.EN-7341

European Commission. (2020). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Chemicals Strategy for Sustainability Towards a Toxic-Free Environment* (COM(2020) 667 final). https://environment.ec.europa.eu/strategy/chemicals-strategy_en

Gocht, T., Berggren, E., Ahr, H. J., Cotgreave, I., Cronin, M. T. D., Daston, G., Hardy, B., Heinzle, E., Hescheler, J., Knight, D. J., Mahony, C., Peschanski, M., Schwarz, M., Thomas, R. S., Verfaillie, C., White, A., & Whelan, M. (2015). The SEURAT-1 approach towards animal free human safety assessment. *ALTEX - Alternatives to animal experimentation*, *32*(1), 9-24. https://doi.org/10.14573/altex.1408041

Haase, A., Barroso, J., Bogni, A., Bremer-Hoffmann, S., Fessard, V., Gutleb, A. C., Mast, J., McVey, E., Mertens, B., Oomen, A. G., Ritz, V., Serchi, T., Siewert, K., Stanco, D., Usmani, S. M., Verleysen, E., Vincentini, O., van der Zande, M., & Cubadda, F. (2024). Proposal for a qualification system for New Approach Methodologies (NAMs) in the food and feed sector: example of implementation for nanomaterial risk assessment. *EFSA Supporting Publications*, *21*(9), 9008E. https://doi.org/10.2903/sp.efsa.2024.EN-9008

Herzler, M., Luijten, M., Marx-Stoelting, P., & Rivière, G. (2025). The European Partnership PARC's role in actively promoting the uptake of new approach methodologies and next-generation risk assessment into regulatory risk assessment practice. *Current Opinion in Toxicology*, *42*, 100517. https://doi.org/https://doi.org/10.1016/j.cotox.2025.100517

ICCVAM. (2018). *A strategic roadmap for establishing new approaches to evaluate the safety of chemicals and medical products in the United States*. https://ntp.niehs.nih.gov/whatwestudy/niceatm/natl-strategy

Kavlock, R. J., Bahadori, T., Barton-Maclaren, T. S., Gwinn, M. R., Rasenberg, M., & Thomas, R. S. (2018). Accelerating the pace of chemical risk assessment. *Chemical Research in Toxicology*, *31*(5), 287-290. https://doi.org/10.1021/acs.chemrestox.7b00339

Leist, M., Tangianu, S., Affourtit, F., Braakhuis, H., Colbourne, J., Cöllen, E., Dreser, N., Escher, S.E., Gardner, I., Hahn, S., Hardy, B., Herzler, M., Islam, B., Kamp, H., Magel, V., Marx-Stoelting, P., Moné, M.J., Lundquist, P., Ottenbros, I., Ouedraogo, G., Pallocca, G., van de Water, B., Vinken, M., White, A., Pastor, M., Luijten, M. (2025). An Alternative Safety Profiling Algorithm (ASPA) to transform next generation risk assessment into a structured and transparent process. ALTEX. https://doi.org/10.14573/altex.2509081.

Marx-Stoelting, P., Rivière, G., Luijten, M., Aiello-Holden, K., Bandow, N., Baken, K., Cañas, A., Castano, A., Denys, S., Fillol, C., Herzler, M., Iavicoli, I., Karakitsios, S., Klanova, J., Kolossa-Gehring, M., Koutsodimou, A., Vicente, J. L., Lynch, I., Namorado, S., . . . Sanders, P. (2023). A walk in the PARC: developing and implementing 21st century chemical risk assessment in Europe. *Archives of Toxicology*, *97*(3), 893-908. https://doi.org/10.1007/s00204-022-03435-7

Moné, M. J., Pallocca, G., Escher, S. E., Exner, T., Herzler, M., Bennekou, S. H., Kamp, H., Kroese, E. D., Leist, M., Steger-Hartmann, T., & van de Water, B. (2020). Setting the stage for next-generation risk assessment with non-animal approaches: the EU-ToxRisk project experience. *Archives of Toxicology*, *94*(10), 3581-3592. https://doi.org/10.1007/s00204-020-02866-4

Mueller, S. O., Dekant, W., Jennings, P., Testai, E., & Bois, F. (2015). Comprehensive summary – Predict-IV: A systems toxicology approach to improve pharmaceutical drug safety testing. *Toxicology in Vitro*, *30*(1, Part A), 4-6. https://doi.org/10.1016/j.tiv.2014.09.016

National Academies. (2007). *Toxicity testing in the 21st century. A vision and a strategy*. The National Academies Press. https://nap.nationalacademies.org/login.php?record_id=11970

OECD. (2017a). *Guidance Document for the use of Adverse Outcome Pathways in developing Integrated Approaches to Testing and Assessment (IATA)* (OECD Series on Testing and Assessment, Issue. OECD Publishing.

OECD. (2017b). *Guidance document on the reporting of Defined Approaches and individual information sources to be used within Integrated Approaches to Testing and Assessment (IATA) for skin sensitisation* (OECD Series on Testing and Assessment, Issue. OECD Publishing.

OECD. (2017c). *Revised guidance document on developing and assessing Adverse Outcome Pathways* (ENV/JM/MONO(2013)6). (OECD Series on Testing and Assessment, Issue. OECD Publishing. https://one.oecd.org/document/env/jm/mono%282013%296/en/pdf

OECD. (2023a). *Guideline No. 497: Defined Approaches on Skin Sensitisation*. OECD Publishing.

OECD. (2023b). *(Q)SAR Assessment Framework: guidance for the regulatory assessment of (quantitative) structure activity relationship models, predictions, and results based on multiple predictions* (ENV/CBC/MONO(2023)32). (OECD Series on Testing and Assessment, Issue. OECD Publishing. https://www.oecd-ilibrary.org/q-sar-assessment-framework-guidance-for-the-regulatory-assessment-of-quantitative-structure-activity-relationship-models-and-predictions_d96118f6-en.pdf

OECD (2023c), Initial Recommendations on Evaluation of Data from the Developmental Neurotoxicity (DNT) In-Vitro Testing Battery, OECD Series on Testing and Assessment, No. 377, OECD Publishing, Paris, https://doi.org/10.1787/91964ef3-en.

PARC. (2023). *Additional Deliverable AD 5.2. Guidance on animal studies in PARC*. https://www.eu-parc.eu/sites/default/files/2023-08/PARC_AD5.2.pdf

Patlewicz, G., & Shah, I. (2023). Towards systematic read-across using Generalised Read-Across (GenRA). *Computational Toxicology*, *25*, 100258. https://doi.org/https://doi.org/10.1016/j.comtox.2022.100258

Paul Friedman, K., Gagne, M., Loo, L.-H., Karamertzanis, P., Netzeva, T., Sobanski, T., Franzosa, J. A., Richard, A. M., Lougee, R. R., Gissi, A., Lee, J.-Y. J., Angrish, M., Dorne, J. L., Foster, S., Raffaele, K., Bahadori, T., Gwinn, M. R., Lambert, J., Whelan, M., . . . Thomas, R. S. (2019). Utility of in vitro bioactivity as a lower bound estimate of *in vivo* adverse effect levels and in risk-based prioritization. *Toxicological Sciences*, *173*(1), 202-225. https://doi.org/10.1093/toxsci/kfz201

Paul Friedman, K., Thomas, R. S., Wambaugh, J. F., Harrill, J. A., Judson, R. S., Shafer, T. J., Williams, A. J., Lee, J.-Y. J., Loo, L.-H., Gagné, M., Long, A. S., Barton-Maclaren, T. S., Whelan, M., Bouhifd, M., Rasenberg, M., Simanainen, U., & Sobanski, T. (2025). Integration of new approach methods for the assessment of data-poor chemicals. *Toxicological Sciences*, *205*(1), 74-105. https://doi.org/10.1093/toxsci/kfaf019

SCCS. (2023). *The SCCS notes of guidance for the testing of cosmetic ingredients and their safety evaluation, 12th revision* (SCCS/1647/22 Corrigendum 2). https://health.ec.europa.eu/document/download/32a999f7-d820-496a-b659-d8c296cc99c1_en?filename=sccs_o_273_final.pdf

Sewell, F., Alexander-White, C., Brescia, S., Currie, R. A., Roberts, R., Roper, C., Vickers, C., Westmoreland, C., & Kimber, I. (2024). New approach methodologies (NAMs): identifying and overcoming hurdles to accelerated adoption. *Toxicological Research*, *13*(2), tfae044. https://doi.org/10.1093/toxres/tfae044

Sunderic, K., Wright, A. M., Kleinstreuer, N., Ledbetter, V., Milora, K. A., Happel, C., Shaughnessy, D., Johnstone, C. P., Casco, D. A., Macioce, N., Marzec, J., Kano, K., Ochocinska, M. J., & Tagle, D. A. (2025). Complement-ARIE: Catalyzing the development and adoption of new approach methodologies. *NAM Journal*, *1*, 100026. https://doi.org/https://doi.org/10.1016/j.namjnl.2025.100026

Thomas, R. S., Bahadori, T., Buckley, T. J., Cowden, J., Deisenroth, C., Dionisio, K. L., Frithsen, J. B., Grulke, C. M., Gwinn, M. R., Harrill, J. A., Higuchi, M., Houck, K. A., Hughes, M. F., Hunter, E. S., III, Isaacs, K. K., Judson, R. S., Knudsen, T. B., Lambert, J. C., Linnenbrink, M., . . . Williams, A. J. (2019). The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. *Toxicological Sciences*, *169*(2), 317-332. https://doi.org/10.1093/toxsci/kfz058

US EPA. (2014). *Next generation risk assessment: incorporation of recent advances In molecular, computational, and systems biology (final report)* (EPA/600/R-14/004). https://ordspub.epa.gov/ords/eims/eimscomm.getfile?p_download_id=520491

Usmani, S. M., Bremer-Hoffmann, S., Cheyns, K., Cubadda, F., Dumit, V. I., Escher, S. E., Fessard, V., Gutleb, A. C., Léger, T., Liu, Y.-C., Mast, J., McVey, E., Mertens, B., Montalvo, D., Oomen, A. G., Ritz, V., Serchi, T., Sieg, H., Siewert, K., . . . Haase, A. (2024). Review of new approach methodologies for application in risk assessment of nanoparticles in the food and feed sector: status and challenges. *EFSA Supporting Publications*, *21*(9), 8826E. https://doi.org/10.2903/sp.efsa.2024.EN-8826

2.2.2. Characterising the protection level and level of confidence associated with next-generation risk assessments

Introduction – the concept of adequate level of protection

A future animal-free Next-Generation Risk Assessment[35] (NGRA) framework should allow to determine whether an "adequate level of protection" (as defined by the regulatory context and problem formulation) for human health and the environment is attained. One of the most frequently heard requirements for NGRA is that, compared to the current system, it needs to provide "the same or a higher level of protection for human health and the environment."

Risk can be defined as "*a measure of the likelihood or probability that damage to life, health, property, and/or the environment will occur as a result of a particular hazard.*"[36]

It is also known from everyday experience that there is no such thing as zero risk in life. Chemical risk management is, therefore, tasked with reducing any risks from chemical exposure to an (albeit very low) residual level that is acceptable from a societal perspective.

In addition, chemical risk management often faces risk-benefit considerations, since the removal of a hazardous chemical from the market might lead to undesirable consequences in other societal areas, e.g. the unavailability of certain products essential for the functioning of society as a whole. Risk management then has to decide about the level of residual risk from uses of that chemical which might still be tolerable in view of the societal benefits.

An "adequate level of protection" can then be defined in general terms as follows:

An **adequate level of protection** represents a level of residual risk for which the combination of severity of effect and expected likelihood of occurrence is acceptable, or at least tolerable, from a societal risk-benefit perspective.

The quantitative benchmarks for "acceptable" or "tolerable" need to be defined on the policy/risk management level. A clear and transparently communicated rationale in this respect will not only allow for more comparable and consistent risk assessments across chemicals and regulatory sectors, but it will also help enhance the acceptance of the resulting risk management measures by society as a whole.

Once such benchmarks have been established, criteria for demonstrating that they have been attained by a given risk assessment need to be provided on the regulatory level. This will then provide the basis for the scientific community to develop and propose methods and workflows within an NGRA framework which are able to demonstrate that these criteria can be met.

However, with respect to the current, mostly animal-based human health chemical risk assessment framework, Chiu & Paoli have noted that:

"*Paradoxically, risk assessments for the majority of chemicals lack any quantitative characterisation as to the likelihood, incidence, or severity of the risks involved.*" (Chiu and Paoli, 2021).

As a consequence, any future NGRA framework is challenged with the "mission impossible" to prove that it lives up to a level of protection that currently is mostly not characterised explicitly for individual assessments, but rather defines safety in an indirect way, e.g. by absence of relevant effects in a set of

[35] In this section, the term "risk assessment" is used collectively to refer to full risk characterisation or its parts, i.e. it includes hazard identification and characterisation as well as external and internal exposure assessment. NGRA then refers to risk assessment strategies that do not yet represent the (present-generation) regulatory standard. In the context of chemical risk assessment it usually refers to strategies using human-centred test methods and decision logics informed by mechanistic understanding.

[36] www.oxfordreference.com, last accessed 2025-07-08

standard tests or by ensuring, via monitoring programmes, that exposure of the human population or the environment is not exceeding "safe" levels.

With one of the world's most sophisticated chemical management systems in place in the EU, the protection level ensured by the current chemical risk assessment framework is very high by comparison to the situation in many other regions of the world. However, this generic statement is of little help for assessing the level of protection present when trying to characterise the level of protection offered e.g. by a reference dose in relation to a specific scenario of exposure to a given hazardous chemical.

Since the current risk assessment framework does not offer a protection level benchmark to which an NGRA result could be compared, there are two strategies for demonstrating that an adequate protection level has been reached:

- a **relative approach,** aiming to compare the protection level represented by NGRA outcomes with that obtained from traditional risk assessments and
- an **absolute approach,** trying to characterise the protection level attained by a given risk assessment, whether NGRA or traditional, in absolute terms.

The long-term roadmap activity described in this section is proposing to develop both strategies in parallel.

Strategies for characterising the protection level provided by risk assessments based on non-animal approaches

"Relative" strategy

Assuming that the protection level provided by the current chemical risk assessment framework is "in general very high," a defensible scientific strategy can be to demonstrate that on average, over a large number of cases with sufficient coverage of chemical and biological space, the output from NGRA, e.g. in the form of a hazard classification result, points of departure (PoDs) for further risk assessment, Predicted No-Effect Concentration (PNEC) in the environment or Health-Based Guidance Value (HBGV), exposure estimates etc., coincides with that from traditional assessments.

A corresponding **relative strategy,** therefore, compares outcomes from traditional animal-based tests with those from NGRA, based on alternatives to animal testing or ultimately on non-animal approaches. To this end, it uses reference datasets to evaluate the similarity in estimated toxicity and associated uncertainties. This approach can be summarised in the following steps:

1. **Characterise the overall uncertainty of the outcomes from NAM-based human health and environmental risk assessments as well as that from traditional approaches.**
   To allow comparison of non-animal approaches-based vs. traditionally derived Points of Departure (PoDs) for hazard characterisation, such as BMDLs, NOAELs/LOAELs, ECx values, first inherent uncertainties of these approaches need to be characterised. In is important to do it at the level suitable to address given hazard concerns as defined in the regulatory system (e.g. allowing a conclusive assessment at the GHS hazard class level).

2. **Provide suitable, well-curated traditional reference datasets and make them broadly available to method developers.**

3. **Establish performance criteria with acceptance by regulators, but also a broad spectrum of stakeholders, which newly developed NGRA strategies would need to fulfil.**
   For example, such criteria could consider uncertainty estimates from step 1. above, e.g. the overall uncertainty associated with traditional PoDs could be used as the basis for defining the required level of concordance between traditional and non-animal approaches-based PoDs). Both the criteria and reference data sets should be broadly disseminated to method developers.

4. **Demonstrate the acceptable performance of newly developed NGRA strategies against the respective reference datasets.**
   To address more complex endpoints and provide risk assessments and regulation with the required level of protection and confidence, non-animal approaches-based solutions will need to combine *in vitro* assays, *in silico* models, knowledge-driven frameworks (e.g., AOPs) and data integration logic. Moreover, since risk is defined as the combination of the severity of effect and the probability of occurrence, probabilistic hazard, exposure and risk assessment methodology will need to be developed and established in support of this approach to safeguard true comparability of traditional and non-animal approaches-based risk assessments.
   Reference datasets should be well-balanced in terms of coverage of the chemical, mechanistic/biological and physicochemical property spaces, as well as biological effect space. At the same time, they must have realistic sample size – for endpoints which are toxicologically more diverse, it should be larger than for narrower ones.

To fully assess the protection level provided by a given risk assessment framework (or a regulatory measure based on it), further elements of the risk assessment process need to be considered, when qualitatively or quantitatively extrapolating from measured or modelled (eco)toxicological PoDs to the real-life exposure situation in humans or the environment. In some of these areas, significant innovation has taken place in recent years, as exemplified by the probabilistic characterisation of PoDs and assessment factors, particularly those addressing population variability, the improvement of physiology-based kinetic (PBK) models or the evolution of human biomonitoring as a tool for characterising internal exposure, all important building blocks of future NGRA frameworks.

A great advantage of the relative approach is that it can be applied immediately, provided an adequate reference dataset is available, cf. (Paul Friedman et al., 2019) or (Lu et al., 2024). Nevertheless, there also inherent problems are associated with it:

- As in the current system, this strategy is not helpful when trying to assess the level of protection provided by a specific individual assessment, e.g. when trying to establish which percentage of an exposed population is protected from which severity of effect and likelihood of occurrence.
- Using the outcome of traditional risk assessments as benchmark propagates possible flaws of the previous system, e.g. the deterministic combination of possibly overconservative worst-case assumptions or animal models with poor predictivity for the target species of the assessment.

“Absolute” strategy

1. **Protection level**

The definition of risk given in the introduction to this section contains two complementary elements – potential damage and likelihood of occurrence. To fully characterise the risk from exposure to hazardous chemicals, the following aspects need to be considered:

- nature and extent or size of a (possible) effect; in chemical risk assessment, the term “endpoint”[37] is frequently used to qualitatively describe the nature of effect whereas “adversity”[38] is used as the synonym for “damage” in the above definition (i.e. not every effect caused by exposure to a chemical).

[37] Note that „endpoint“ in this sense may refer to different levels of biological organisation, i.e. molecular, cellular, organ, system, organism, population, ecosystem or even the planetary level.
[38] Cf. the definition of "adverse effect” in (WHO IPCS, 2004):“Change in the morphology, physiology, growth, development, reproduction, or life span of an organism, system, or (sub)population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influences”.

- likelihood of the occurrence of an adverse effect (over time depending on the nature of the population or system exposed and the variability in exposure level, duration and frequency as well as susceptibility (sensitivity and vulnerability)[39] of that population or system.

As the term "likelihood of occurrence" implies, the protection level provided by an individual risk assessment can be characterised using probabilistic assessment methodologies. Maertens and co-workers have recently described the benefits of such approaches over current "deterministic"[40] risk assessment outputs (Maertens et al., 2024; Maertens et al., 2022; Maertens et al., 2025). EU Horizon 2020 projects such as RISK-HUNT3R (Moné et al., 2020) or ONTOX (Vinken et al., 2021) or the European Partnership for the Assessment of Risks from Chemicals (PARC, Herzler et al. (2025))[41] are currently working on the integration of probabilistic methodology into NGRA frameworks and workflows.

It may also be argued that probabilistic methods more adequately reflect the reality of potential chemical effects on populations. A conceptual framework has been provided by Chiu and Slob (2015) and further elaborated on by WHO IPCS (2018), who also developed a tool for practical application of "Approximate Probabilistic" risk characterisation, APROBA(-Plus)[42].

However, experience with the practical application of such methodology is currently limited (cf. e.g. EFSA (2019) or BfR (2023)) and required input data may be lacking or difficult to access. Alternatives to animal approaches, and in particular high-throughput methods, may be particularly suited to generate such data, since e.g. more concentrations or replicates could be tested at lower cost compared to adding further animals. In addition, alternatives to animal approaches can be utilised to assess e.g. intraspecies variability (Rusyn et al., 2022; Zeise et al., 2013)

Consequently, the envisaged **absolute strategy** aims to develop an overall conceptual approach for probabilistic characterisation of human risk, integrating data on exposure, effect severity, and population sensitivity and vulnerability. This method expresses risk as a probability distribution, explicitly incorporating uncertainty (see section below) and variability. Probabilistic risk metrics – e.g., 95th percentile thresholds – can be used to standardise protection levels across different regulatory applications. This approach ultimately allows for a transparent and standardised assessment of the protection level. Compared to the relative approach, it can also be applied – once established – in situations with sparse data or where PoDs are not directly available.

2. Level of confidence

Developing robust concepts and methodology for the probabilistic characterisation of exposure, hazard or risk is an important step, but not enough in itself. In addition, an adequate level of confidence (as defined by the problem formulation and regulatory context) needs to be attained, especially when concluding on the absence of relevant hazard, exposure or risk.

Approaches to achieve confidence in NGRA frameworks, non-animal approaches and results based on them in general terms, e.g. by validation, are addressed in other sections of this document. While a prerequisite, this trust alone, however, does not help when trying to assess the confidence that one can place in a specific risk assessment outcome.

---

[39] In line with WHO IPCS (2018), more "sensitive" individuals in a population show the same effect size at a lower exposure than less sensitive ones; more "vulnerable" individuals are more strongly affected by the same effect size.

[40] In this context, deterministic refers to the unambiguous relationship between e.g. an HBGV, given as a single number, and the protection level of an exposed individual. In contrast, a probabilistic risk assessment describes this relationship in terms of a probability distribution, bearing witness to the fact that e.g. different individuals may possess different susceptibility to a chemical's adverse effect.

[41] EU Horizon 2020 projects: RISK assessment of chemicals integrating HUman centric Next generation Testing strategies promoting the 3Rs | RISK-HUNT3R | Project | Fact Sheet | H2020 | CORDIS | European Commission, Ontology-driven and artificial intelligence-based repeated dose toxicity testing of chemicals for next generation risk assessment | ONTOX | Project | Fact Sheet | H2020 | CORDIS | European Commission, Partnership for the Assessment of Risks from Chemicals | PARC | Project | Fact Sheet | HORIZON | CORDIS | European Commission.

[42] https://www.rivm.nl/en/aproba-plus (accessed 2024-10-04)

All methodology used in the context of chemical risk assessment either represents models (*in silico*, *in vitro*, *in vivo*) or measurements (subject to experimental error) with limited sample size. Both are then used to make inferences about larger collectives (of chemicals, populations or species). Without exception, therefore, all chemical risk assessments are associated with uncertainty.

Uncertainty has been defined by WHO IPCS as

"*imperfect knowledge concerning the present or future state of an organism, system, or (sub)population under consideration.*" (WHO IPCS, 2004)

Acc. to the WHO IPCS guidance document on characterising uncertainty in hazard characterisations,

"*it can be further defined as lack of knowledge regarding the "true" value of a quantity, lack of knowledge regarding which of several alternative model representations best describes a system of interest, or lack of knowledge regarding which probability distribution function and its specification should represent a quantity of interest.*" (WHO IPCS, 2018)

In contrast, according to the same document, **variability** marks the

"*heterogeneity of values over time, space or different members of a population, including stochastic variability and controllable variability. Variability implies real differences among members of that population. For example, different individual persons have different intake and susceptibility. In relation to human exposure assessment, differences over time for a given individual are referred to as intraindividual variability; differences over members of a population at a given time are referred to as interindividual variability.*" (WHO IPCS, 2018)

Biological variability is a reality that may be measured and – e.g. in terms of experimental design – partly controlled but not changed. In contrast, uncertainties (e.g. about the variability of the response of a given population towards a chemical stressor) may in principle be remediated by generation of further knowledge, e.g. by applying a refined assessment strategy or by generating additional data.

Under current chemical risk assessment frameworks, assessment outcomes are mostly only reported in a deterministic[43] way, without a transparent, quantitative statement regarding the associated uncertainties. This is often compensated by using what are supposed to be worst-case or justified generic assumptions, the combination of which, however, may result in highly conservative (and potentially overconservative) assessments.[44] Notably, such (over–) conservatism may not be a problem or even be desired in certain cases but needs to be made at least transparent to avoid misinformed risk management decisions. An additional, undesirable consequence of this practice is that different risk assessments may differ vastly with respect to their inherent uncertainties, rendering their results incomparable.

A non-exhaustive list of elements in current chemical risk assessment typically associated with uncertainties includes

- the outcome of dose-response analyses,
- extrapolations across species, exposure routes or durations,
- assumed use patterns for chemicals/products, or
- the variability with respect to susceptibility towards a certain chemical's toxicity within the target population or between target species,

[43] For example, a health-based guidance value (HBGV, single number) is compared to an exposure estimate (single number), resulting in a single number (e.g. the risk characterisation ratio, RCR = exposure/HBGV) to characterise a possible risk.
[44] For example, the likelihood that, for a given substance, the discrepancy between the sensitivity of rats vs. humans towards that chemical's toxicity is tenfold AND a person's individual susceptibility to the effect is again tenfold that of the population median AND exposure of that person is at the maximum level experienced in the population is extremely low, yet, such a combined assumption is frequently made in CRA.

- exposure assessments, as use tonnages, uses, associated use patterns, emissions, fate & leaching potential are often no/rarely published or known.

In NGRA, some of these aspects, e.g. the extrapolation across species in human health assessments, will become less relevant. However, additional uncertainties will have to be considered, e.g. in relation to extrapolating the results from *in vitro* experiments to the *in vivo* situation (in-vitro-to-in-vivo extrapolation, IVIVE).

Qualitative statements to the end that uncertainties are present are of limited practical help in risk management and communication. The same holds for semi-quantitative statements that uncertainty was "low", "medium" or "high", as long as no exact boundaries between these classes are provided (but then these boundaries could also be stated directly). In addition, acceptance of an assessment outcome implies that, whatever the uncertainty, it was considered acceptable in the end.

The benefits of quantitative approaches to uncertainty assessment have again been described by WHO IPCS:

"*More transparently representing in quantitative form the confidence we can have in toxicological risk projections and estimates of the relationship between dose and health effect, thereby facilitating choices of preventive measures and/or further information gathering by risk managers. For instance, health-based guidance values […] may be defined based on a pre-specified and harmonised level of conservatism, or estimated health risks for a given exposure situation can be expressed in terms of an uncertainty distribution or a confidence interval.*" (WHO IPCS, 2018)

Another noteworthy aspect of uncertainty in the context of chemical risk assessment is that it works in both directions, i.e. a statement on the risk posed by a certain exposure scenario is uncertain regarding a potential under- as well as overestimation of hazard, exposure and/or risk.

The degree of uncertainty and confidence acceptable under a given problem formulation and regulatory context needs to be defined at the policy level and concrete benchmarks for acceptance need to be set at the regulatory level. The ability to provide a transparent measure of the uncertainties inherent in a risk assessment would then also mark a superior quality of NGRA over the current chemical risk assessment framework, where such a measure is often not available or provided. Where uncertainties and confidence cannot be sufficiently quantified, (semi)qualitative methods may be applied, but this will result in additional uncertainty about the outcome of the assessment, which should be made transparent.

3. Challenges and aspects to be considered under the "absolute" approach

A non-exhaustive selection of specific challenges and aspects to be considered when developing strategies demonstrating the level of protection and increasing the level of confidence with NGRA based on non-animal approaches has been compiled by the PARC project.[45]

Conclusion

Since the absolute strategy relies on input from the relative approach, both methods are complementary and will be developed concurrently. The relative approach could be used immediately where appropriate reference datasets exist or can be curated within a reasonable period. In contrast, the absolute strategy requires initial consensus on the most effective conceptual framework before proceeding.

Nevertheless, introducing such concepts requires substantial preparatory work: defining reference datasets, characterising uncertainty in both traditional and non-animal method-based estimates and agreeing on performance benchmarks. Therefore, working on developing and introducing strategies for characterising the protection level provided by risk assessments with non-animal approaches will be a long-term project, which will need to be supported by collaborations like PARC and EU research projects and involving broad participation of regulators and stakeholders. Overall, the aim is to provide a sound scientific basis for chemical risk assessment, management and communication, as well as practical methodological approaches for characterising uncertainty, variability, and protection level. In

[45] Source: PARC Deliverable AD2.1, submitted to the European Commission for approval, draft version available at https://www.parcopedia.eu/wp-content/uploads/2024/10/PARC_AD2.1_submitted_approval_pending.pdf.

this way, it will help to achieve regulatory and societal acceptance of non-animal approaches by making the concept of "protection" both scientifically tangible and transparently communicable.

References

BfR (2023): Bisphenol A: BfR proposes health based guidance value, current exposure data are needed for a full risk assessment : BfR Opinion no. 018/2023, issued 19 April 2023. DOI: 10.17590/20230419-114234-0

Chiu W.A. and Paoli G.M. (2021): Recent advances in probabilistic dose–response assessment to inform risk-based decision making. Risk Analysis 41 (4), 596 - 609. DOI: 10.1111/risa.13595

Chiu W.A. and Slob W. (2015): A unified probabilistic framework for dose–response assessment of human health effects. Environmental Health Perspectives 123 (12), 1241 - 1254. DOI: 10.1289/ehp.1409385

EFSA (2019): Establishment of cumulative assessment groups of pesticides for their effects on the thyroid. EFSA Journal 17 (9), e05801. DOI: 10.2903/j.efsa.2019.5801

Herzler M., Luijten M., Marx-Stoelting P., and Rivière G. (2025): The European Partnership PARC's role in actively promoting the uptake of new approach methodologies and next-generation risk assessment into regulatory risk assessment practice. Current Opinion in Toxicology 42, 100517. DOI: 10.1016/j.cotox.2025.100517

Lu E.-H., Ford L.C., Chen Z., Burnett S.D., Rusyn I., and Chiu W.A. (2024): Evaluating scientific confidence in the concordance of in vitro and *in vivo* protective points of departure. Regulatory Toxicology and Pharmacology 148, 105596. DOI: 10.1016/j.yrtph.2024.105596

Maertens A., Antignac E., Benfenati E., Bloch D., Fritsche E., Hoffmann S., Jaworska J., Loizou G., McNally K., Piechota P., Roggen E.L., Teunis M., and Hartung T. (2024): The probable future of toxicology – probabilistic risk assessment. ALTEX - Alternatives to animal experimentation 41 (2), 273-281. DOI: 10.14573/altex.2310301

Maertens A., Golden E., Luechtefeld T.H., Hoffmann S., Tsaioun K., and Hartung T. (2022): Probabilistic risk assessment – the keystone for the future of toxicology. ALTEX - Alternatives to animal experimentation 39 (1), 3-29. DOI: 10.14573/altex.2201081

Maertens A., Kincaid B., Bridgeford E., Brochot C., de Carvalho e Silva A., Dorne J.-L.C.M., Geris L., Husøy T., Kleinstreuer N., Ladeira L.C.M., Middleton A., Reynolds J., Rodriguez B., Roggen E.L., Russo G., Thayer K., and Hartung T. (2025): From cellular perturbation to probabilistic risk assessments. ALTEX - Alternatives to animal experimentation. DOI: 10.14573/altex.2501291 (last accessed 2025/06/27)

Moné M.J., Pallocca G., Escher S.E., Exner T., Herzler M., Bennekou S.H., Kamp H., Kroese E.D., Leist M., Steger-Hartmann T., and van de Water B. (2020): Setting the stage for next-generation risk assessment with non-animal approaches: the EU-ToxRisk project experience. Archives of Toxicology 94 (10), 3581-3592. DOI: 10.1007/s00204-020-02866-4

Paul Friedman K., Gagne M., Loo L.-H., Karamertzanis P., Netzeva T., Sobanski T., Franzosa J.A., Richard A.M., Lougee R.R., Gissi A., Lee J.-Y.J., Angrish M., Dorne J.L., Foster S., Raffaele K., Bahadori T., Gwinn M.R., Lambert J., Whelan M., Rasenberg M., Barton-Maclaren T., and Thomas R.S. (2019): Utility of in vitro bioactivity as a lower bound estimate of *in vivo* adverse effect levels and in risk-based prioritization. Toxicological Sciences 173 (1), 202-225. DOI: 10.1093/toxsci/kfz201

Rusyn I., Chiu W.A., and Wright F.A. (2022): Model systems and organisms for addressing inter- and intra-species variability in risk assessment. Regulatory Toxicology and Pharmacology 132, 105197. DOI: 10.1016/j.yrtph.2022.105197

Vinken M., Benfenati E., Busquet F., Castell J., Clevert D.-A., de Kok T.M., Dirven H., Fritsche E., Geris L., Gozalbes R., Hartung T., Jennen D., Jover R., Kandarova H., Kramer N., Krul C., Luechtefeld T., Masereeuw R., Roggen E., Schaller S., Vanhaecke T., Yang C., and Piersma A.H. (2021): Safer chemicals using less animals: kick-off of the European ONTOX project. Toxicology 458, 152846. DOI: 10.1016/j.tox.2021.152846

WHO IPCS (2004): IPCS Risk Assessment Terminology. International Programme on Chemical Safety. World Health Organization, Geneva. https://www.who.int/publications/i/item/9241562676 (last accessed 2024-09-04)

WHO IPCS (2018): Guidance on evaluating and expressing uncertainty in hazard assessment. International Programme on Chemical Safety. World Health Organization, Geneva. https://www.who.int/publications/i/item/9789241513548 (last accessed 2024-09-04)

Zeise L., Bois F.Y., Chiu W.A., Hattis D., Rusyn I., and Guyton K.Z. (2013): Addressing human variability in next-generation human health risk assessments of environmental chemicals. Environmental Health Perspectives 121 (1), 23-31. DOI: 10.1289/ehp.1205687

2.2.3. Hazard classification for systemic toxicity based on non-animal data (EPAA Designathon)

Rationale and background

This chapter reflects on possibilities for hazard classification for human systemic toxicity solely based on the use of alternatives to animal testing (or new approach methodologies, NAMs). This involves the design of a potential future classification scheme (Next Generation Classification and Labelling) ensuring equivalent protection, by calibrating the system with substances that are currently classified.

		Activity (NAM-based toxicodynamics)		
		High	Medium	Low
Potential Systemic Availability (NAM-based toxicokinetics, based on ADME properties)	High	H	H	M
	Medium	H	M	L
	Low	M	L	L

Figure 1. Proposal for potential new classification scheme for chemicals based on three levels of concern (High, Medium and Low) according to Berggren E & Worth AP (2023).

A proof-of-concept of this idea is the focus of the EPAA Designathon project, which started in 2023[46]. Central to this project is a level of concern matrix (Figure 1). The idea is that results from eventual non-animal approaches-based solutions should fall somewhere on the matrix, reflecting different levels of concern (high, medium, low) which would be the basis for different risk management measures (Berggren and Worth, 2023; Worth and Berggren, 2025; Worth et al., 2025). The two dimensions of the matrix refer to intrinsic properties, one axis being information on toxicokinetics (TK) and ADME properties and the other on toxicodynamics (TD), both based on non-animal approaches.

Moreover, the project also explores the possibility for a systemic toxicity class based on single alternative methods. This might be feasible since adverse outcomes are not mutually independent from a mechanistic point of view – a limited number of upstream key events are likely to underpin a spectrum of adverse outcomes due to divergent AOP networks (Figure 2).

[46] EPAA designathon challenge. https://single-market-economy.ec.europa.eu/calls-expression-interest/epaa-designathon-human-systemic-toxicity_en

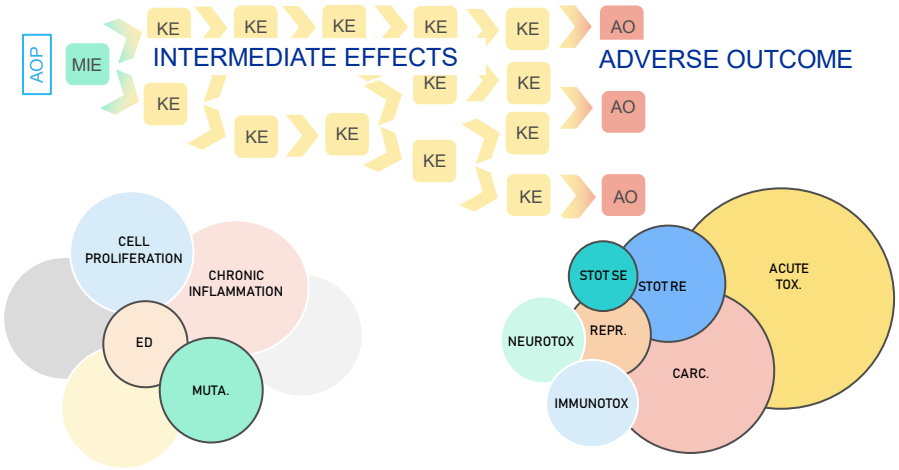


Figure 2. Hypothesis that adverse outcomes associated with systemic toxicity are underpinned by a limited number of upstream key events.

Further considerations will be necessary as part of the roadmap implementation phase on how the new hazard class could be brought into operation. For instance, it could be beneficial to operate for an interim phase the current GHS system for classification and the future GHS hazard class in parallel (Figure 3). This would on one hand ensure continuity with the current GHS system, while both systems could mutually inform each other. Whether to introduce such options after the development phase, requires careful reflection, including on operability and appropriateness. It will further require consultations of EU regulators and international partners. In addition, anticipating the need to protect for systemic health concerns not yet considered (e.g. immunotoxicity, endocrine disruption), the new system is designed to embrace any additional endpoint, while the current system tends to establish either new classes or define additional criteria that are included in current classes.

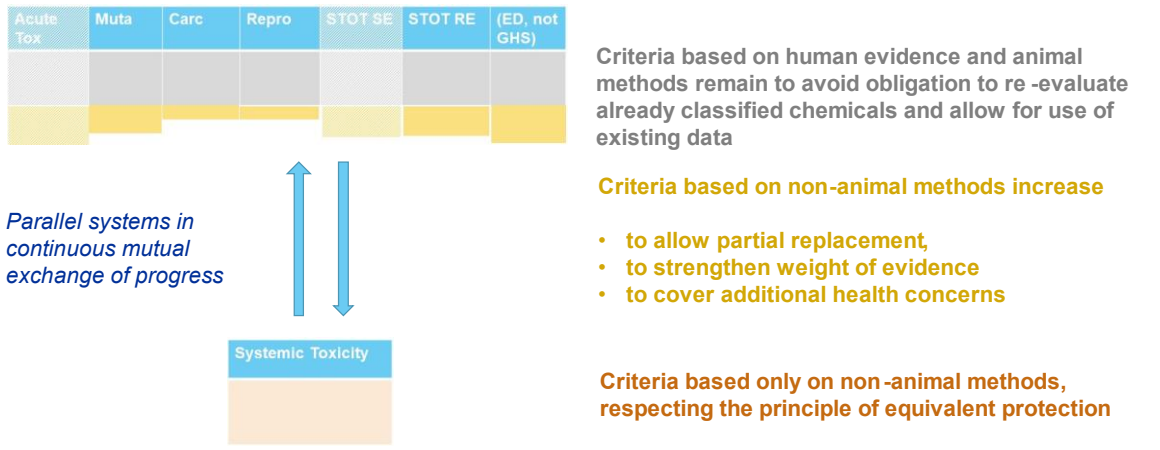


Figure 3. Co-existence of a new NAM-based systemic toxicity class with existing hazard classes

For other toxicological effects development of a similar classification strategy could be explored.

References

Berggren E & Worth AP (2023). Towards a future regulatory framework for chemicals in the European Union — Chemicals 2.0. Regulatory Toxicology & Pharmacology 142: 105431.

Worth AP & Berggren E (2025). A twin transition in regulatory toxicology: moving toward Chemicals 2.0 and phasing out animal testing. Toxicological Sciences 203: 160–165.

Worth AP, Berggren E & Prieto P (2025). Chemicals 2.0 and Why We Need to Bypass the Gold Standard in Regulatory Toxicology. Alternatives to Laboratory Animals 53: 21-25.

2.2.4. Long-term change - A new paradigm for regulatory environmental safety assessments

Introduction and recommendations

Discussions with regulators and stakeholders during the development of the roadmap made it evident that replacing animal testing with one-to-one with equivalent non-animal approaches can only be achieved in specific and limited number of cases. Non-animal approaches provide information that is different from the animal tests that are currently used. Information from non-animal approaches is well suited to provide the information necessary to reach the policy goals of protecting human health and the environment. However, regulatory assessments, and in the end also chemical legislation, has to be re-designed and tailored to the information that non-animal approaches provide if a transition should succeed. A paradigm shift is needed for regulatory environmental safety assessments (ESA) that facilitates incorporation of non-animal approaches while guaranteeing a high level of environmental protection.

To achieve the paradigm shift, it is recommended to develop and implement the use of animal-free exposure-led hypothesis-driven environmental Next Generation Risk Assessment (NGRA) in the regulatory context. The new system should include considerations of how to define a system for classification and labelling that relies on innovative non-animal approaches only but offers at least the same level of protection. The new system based on NGRA also needs to provide the information required for hazard characterisation and the derivation of thresholds.

Reaching such a system requires the following implementation and research activities:

- Development of a generic bioactivity battery covering the most relevant biological pathways to be used for the hazard identification of chemicals with unknown modes of action.
- Development of proposals for prioritising biological pathways, expected consequences, taxa and traits with expected higher risk.
- An understanding of how to use information from sub-organism level to extrapolate to the population level or higher.
- Designing and implementing cross-sector case studies for hazard identification and characterisation considering two parallel approaches, one using exclusively non-animal approaches, the second combining non-animal approaches and invertebrate testing.
- Developing tiered, hypothesis-driven strategies for the use of non-animal approaches for hazard identification and characterisation.
- Integrating information on distribution and fate of chemicals to upgrade the hazard characterisation case studies as exposure-led environmental NGRA assessments.
- Development and implementation of computational tools to facilitate gathering, organising and interpreting of anticipated datasets e.g., according to FAIR principles.
- Exploring the links to human health NGRA tools/data where applicable under a one-health concept.

The above points require to extend the knowledge in multiple areas relevant for ESA, and *inter alia* the following:

- Extending the biological and ecological knowledge that is integrated into ESAs, e.g., natural variability in impact on risk drivers due to life history traits of relevant organisms under different ecological scenarios and conditions;
- Extending the chemical and non-chemical stressor knowledge integrated into ESAs, e.g., spatial and temporal exposure patterns, multiple stressors on ecosystem services;
- Leveraging computational advances to build tools and models to support the gathering, integration and interpretation of exposure datasets, as well as biological and ecological knowledge, for the systems-based risk assessments, e.g., digital twins and AI advances.

The current regulatory system for environmental safety assessment

The current ESA paradigm was developed around the 1980s. Since then, the knowledge on ecological stress and capacities for predicting the impacts of chemical exposure on organisms, populations or ecosystems has been significantly improved.

A main advantage of the current paradigm regarding regulatory use is the high level of predictability. The combination of a) preestablished standard study protocols conducted on selected species and endpoints, and b) agreed extrapolation factors, limits the needs for applying expert judgement to those cases with quality issues on the studies and the confirmation that the effect is associated to the exposure (e.g. unclear dose-response relationship).

An advantage from the regulatory perspective is the possibility for conducting an ESA with a minimum set of data, e.g. under REACH for some substances, it is possible to derive PNECs for water, sediment and soil using just three aquatic short-term studies or acute toxicity QSAR estimations, if their reliability is demonstrated. This, however, does not apply to poorly water-soluble substances. The flip side of this coin is the limited relevance of the derived toxicity values, particularly for addressing delayed effects (i.e. effects seen after exposure has ended), at the ecosystem level, and long-term consequences at the population and/or community level.

The methodological limitations can be grouped in the following aspects:

- Coverage of species, taxonomic groups and traits: e.g. a cladoceran crustacean represents all aquatic invertebrates.
- Coverage of ecologically relevant endpoints: e.g. immunotoxicity, sublethal morbidity, stress adaptation, etc. are not considered.
- Coverage of biological mechanistic differences for the selected endpoints: e.g. reproduction effects linked to the reproductive approach of the selected species such as parthenogenicity for daphnids.
- Coverage of ecological conditions: laboratory assays conducted under optimised development conditions for the test species.
- Coverage of ecological interactions: single species laboratory test with no competition, etc.

The conceptual limitations are mostly linked to the current regulatory approach, the first two in the list below are already highlighted in the EU Chemicals Strategy for Sustainability, which calls for the consideration of aggregate and combined exposures:

- Lack of integration of exposures to the same substance from different uses in different sectors (and in some cases also within the same sector, e.g. due to different applicants or to assessing each use individually in the case of pesticides and biocides).
- Lack of consideration of the combined effects linked to simultaneous or subsequent exposures to different substances.
- Focus on direct toxic effects, do not consider that indirect effects not identified by apical endpoints, such as increased susceptibility to diseases or behavioural changes, may be more relevant for population and community dynamics.
- Overgeneralisation and insufficient consideration of real environmental and ecological variability within the EU.
- Approach, based on setting thresholds for each individual substance, puts the focus on the regulatory decision, but does not provide information on the actual level of risk associated to the decision.
- There is no information related to the uncertainty associated to the ecotoxicological threshold.

Due to the above-mentioned limitations, the current approach based on laboratory studies does not inform on:

- Type, likelihood and magnitude of the potential environmental effects linked to the final regulatory decision,
- Influence of the variability of biotic (ecosystem structure and functions), and abiotic factors (including climate and landscape distribution) on the environmental risks characterisation,

- Possibility of indirect environmental effects that may affect population and community dynamics and impact on biodiversity.

The current regulatory environmental safety assessment includes several steps: hazard identification, hazard characterisation that quantitatively assesses the magnitude of adverse effects (thresholds), and risk assessment, which combines hazard identification, hazard characterisation with information on fate and exposure of substances in the environment.

Conceptually, the concept used in ESA for hazard characterisation and risk assessment is an adaptation of the system for human health assessments, based on the extrapolation of the no-effect level using generic extrapolation factors. The main modification is in the selection of the toxicity endpoints. The approach for environmental risk assessment (ERA) is to consider only those effects that are assumed to be relevant at population level, in particular survival, growth and reproduction. As a consequence, the assumed "no-effect level" from the most sensitive species used as point of departure may in fact produce significant effects at individual level as long as no statistically significant effects are observed on survival, growth and reproduction.

The current extrapolation factors are a legacy from risk management decisions in the second half of the 20th Century (Tarazona 1997), lacking a sound scientific basis (Chapman et al. 1998; Smith and Cairns 1993). In addition, species sensitivity is usually distributed along several orders of magnitude (Newman et al. 2000), indicating that real interspecies variability is much larger than that covered by the interspecies extrapolation factors, thus the real level of protection depends on the availability or not of toxicity data for the most sensitive taxonomic groups.

The current risk assessment paradigm includes parallel quantitative assessment for hazard and exposure, including environmental fate, which are then compared using their quotient to characterise the risk. The most common approach for hazard characterisation is to set an acceptability threshold for each environmental compartment or exposure pathway, i.e. a predicted-no-effect-concentration (PNEC). PNECs are derived by applying an assessment factor (to cover intra/interspecies and lab-to field variabilities) to the lowest level of effect or no effect, i.e. lowest relevant lethal (or effect) concentration that effects 50% or 10% of the animals ($L(E)C_{50}$/$L(E)C_{10}$) or no-observed-effect concentration (NOEC), obtained in experimental studies from pre-established standard species.

Differences between legislative areas exist. In some areas, for instance plant protection products (PPPs), the approach is modified. The acceptability threshold is established for each taxonomic group and exposure timeline (i.e. acute and chronic). Exceedance of this threshold requires further refinement or assumption of unacceptable risk. The problem formulation includes different exposure-to-target lines according to the use patterns.

Under REACH, the main exposure-to-target lines are for the aquatic compartment (including sediment), soil, and predators exposed via the aquatic and terrestrial food chains. For PPPs, additional assessment lines are included for non-target terrestrial vertebrate and invertebrate groups.

The interspecies assessment (or extrapolation) factors for PNEC derivation under REACH depend on the number of taxonomic groups with available information and, when it is assumed that steady-state can be achieved in acute tests, an additional acute to chronic extrapolation factor may be applied when only acute toxicity information is available. For PPP, the uniform principles provide acceptability thresholds, which conceptually are equivalent to the overall extrapolation factor, for each assessed taxa and exposure time. Species sensitivity distributions (SSDs) and higher-tier studies can be used for refinement.

There are no solid scientific bases for setting the interspecies extrapolation factors, and it should be noted that the final risk acceptability level is also heavily dependent on the exposure scenarios developed in the guidance documents. For example, the PECwater for industrial and domestic emissions is estimated after the mixing zone, using a dilution factor, thus the selection of the dilution factor is critical for understanding the overall level of protection.

Opportunities for a paradigm shift

The short- and medium-term approaches in the roadmap towards phasing out animal testing for chemical safety assessment focus on the replacement of animal testing (or reduction or refinement) with methods or a combination of methods that provide the same or similar information as the current animal testing.

The long-term option addressed in this document explores the needs for a paradigm shift to develop non-animal ESA approaches, addressing, in addition, the limitations of the current system, through the use of recent scientific developments and novel approaches for supporting regulatory decisions.

The paradigm shift requires to explore the capacity of mechanistic information from non-animal approaches to address the methodological and conceptual limitations listed above. The proposed methodology is based on exposure-led, hypothesis-driven, environmental next generation risk assessment methodologies using non-animal approaches.

With a focus on non-animal approaches, environmental NGRA will provide more informative risk characterisations of the assessed chemical, i.e., a risk profile identifying the biological pathways expected to be affected at the estimated exposure levels and the taxa, traits and functions with higher risk, including the variability associated to different environmental conditions in the exposure assessment, and ecological considerations for the extrapolation of a hazard profiling based on non-animal approaches.

The basis of the current paradigm is the extrapolation of effects from individual organisms to populations. The additional complexity for NGRA based on non-animal approaches is represented by the need to explore effects at the lower levels of biological organisation and link them, to assess impact at much higher levels, as depicted in Figure 4. The development of tools that enable to go beyond the impact on the single organism by extrapolating to population and community level should be a priority, as the aim of environmental NGRA should be increasing the environmental relevance, beside moving away from animal testing.

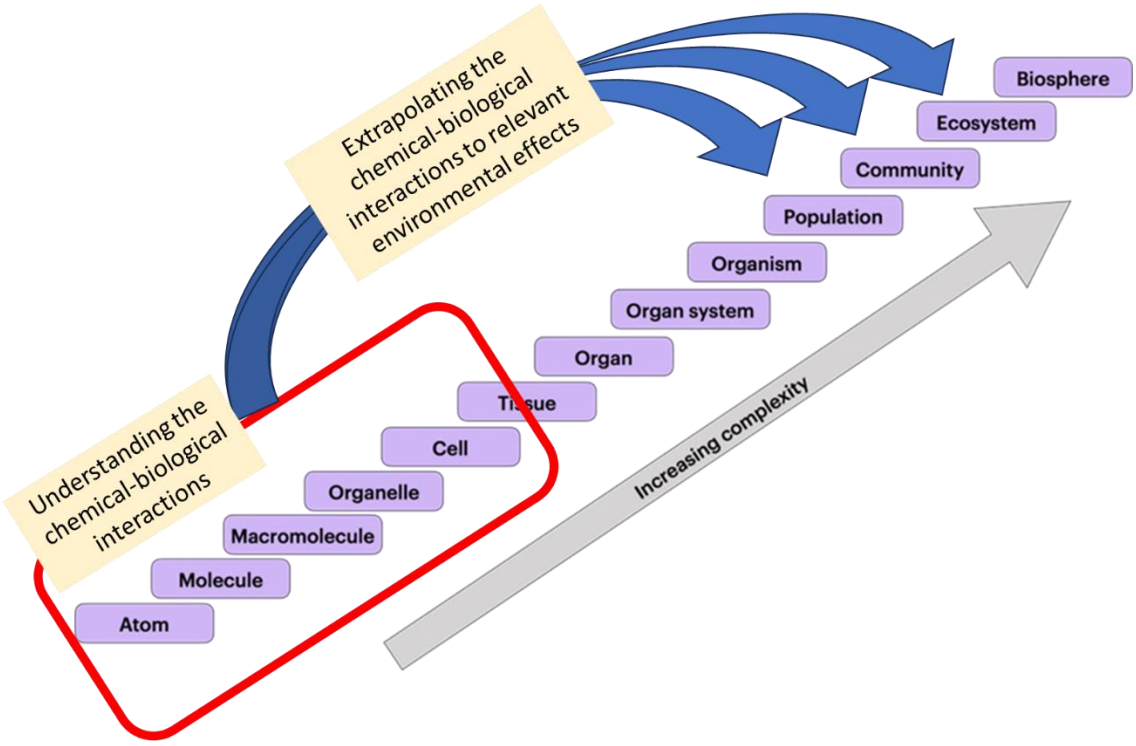


Figure 4: The complexity line for NAM-based NGRA environmental assessments[47]

[47] Source of diagram: M. Hardy, College Biology I, modified.

The exposure-led and the hypothesis-driven NGRA is linked to more flexible problem formulations, replacing fixed information requirements by Integrated Approaches to Testing and Assessment (IATA). Following the IATA process, the problem formulation starts with the assessment of existing information and the identification of which information is needed for regulatory decision-making. Exposure considerations should lead the assessment, and this represents a clear benefit in the EPAA context as emissions and most environmental fate processes can be addressed without the use of animal methods. In the implementation, it should be considered that the possibility for targeting the exposure assessment is much higher in some sectors than in others, requiring adaptations to the specific regulatory frameworks. The focus of the exposure assessment should be to identify which populations and communities could be exposed and under which circumstances, targeting the identification of the information needs to cover the effect assessment.

Several NAM-based approaches are under development to cover the effect assessment and, in general, these are linked to Adverse Outcome Pathway (AOP) frameworks (Brooks et al. 2023). The "offer" includes advanced molecular methods (Schumann et al. 2024b), effects at genetic level and the associated pathways (Rivetti et al. 2023), and high throughput *in vitro* bioactivity (Schaupp et al. 2023); frequently supported by computational methods (Kosnik et al. 2024; Schumann et al. 2024a). The assessment of interspecies similarities and differences is essential in all cases, and evolutionary conservation represents a key element (Dufourcq Sekatcheff et al. 2024; Rivetti et al. 2023).

These different options are complementary, and the new paradigm should allow the development and implementation of hypotheses supported by any of the options in a weight of evidence approach. Phylogenetic studies indicate that, in general, biological pathways are widely shared among organisms from different taxa, however, the same pathway may be associated to different physiological processes, consequently the impairment may result in very different adverse outcomes; e.g., the alteration of the thyroid pathway has been associated with neurodevelopmental effects in mammals (AOP 42), metamorphosis impairment in amphibians (AOP 176), and swimming alteration in fish (AOP 176)[48]. This situation facilitates the screening process for setting a battery of relevant bioactivities but requires sufficient knowledge on the biology and ecology of the different species in order to assess the possible outcomes and identify the risk drivers in terms of integration of exposure, biological role and ecological traits.

In addition to cross-species extrapolation, improving ecological relevance requires a new brainstorming to link current population drivers (survival, reproduction, growth and development) with species ecology and environmental conditions, as well as to consider other endpoints, such as behavioural effects or disease susceptibility. The current paradigm is based on the consideration of the "population relevance" of the endpoints measured *in vivo* studies, but population relevance is not unequivocally defined, and the relative relevance of each endpoint depends on the species, trait, environmental conditions and ecological status. For example, behavioural alterations or increased susceptibility to infectious diseases due to immunosuppression, may be much more relevant in some circumstances than effects on survival or reproduction. As a consequence, a well-designed bioactivity battery linked to sufficient knowledge on the biology and ecology of wildlife may be much more informative in terms of potential ecological impacts than observational laboratory animal studies in a limited number of species.

There are several activities already addressing this issue, e.g., Partnership for the Assessment of Risks from Chemicals | Parc (eu-parc.eu), and collaboration is essential. Some proposals keep exposure and effects as independent assessments (Rivetti and Campos 2023); others suggest a more integrative approach of the exposure and effect assessments (Di Nicola et al. 2023).

The proposed implementation is summarised in Figure 5. In addition to the integration of exposure and effects, a key element is the consideration of the environmental characteristics and the ecological conditions, which obviously are interlinked.

[48] https://aopwiki.org/ accessed on 27 Feb 2025

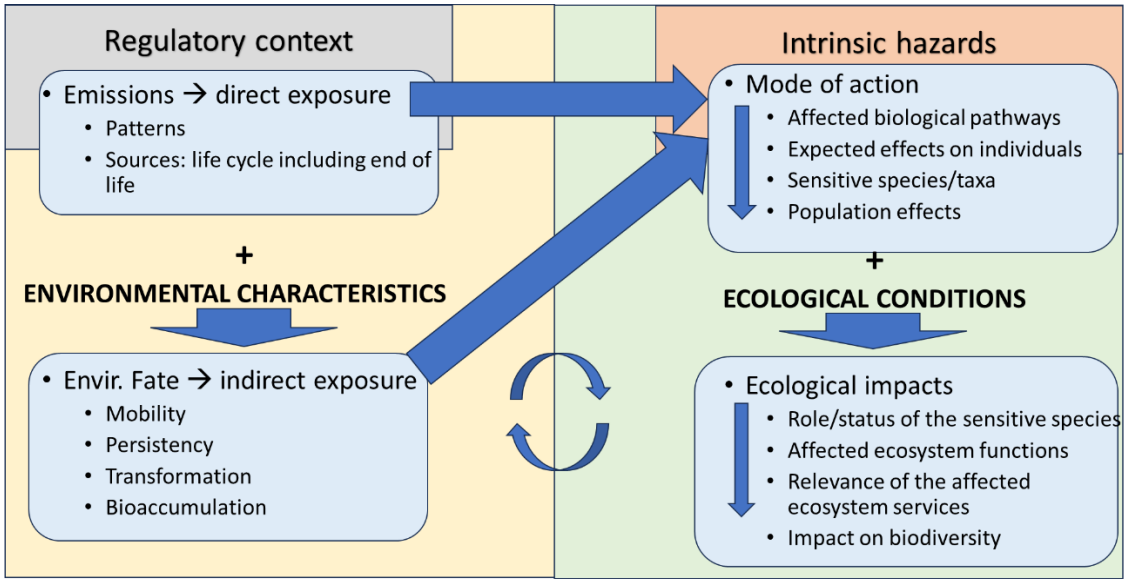


Figure 5. Proposed implementation of environmental NGRA.

In the regulatory context, in addition to exposure-led risk assessments, the identification of intrinsic hazards not linked to exposure assessments is also relevant, for example CLP aims to cover hazards of non-intended emissions as a result of accidents or spills. Under REACH, the hazard identification triggers exposure assessment and risk characterisation when certain hazard criteria are met. The environmental and ecological conditions are also relevant, for example for assessing the hazards of accidental non-intended emissions. The current technologies offer options for adapting the information relevant for different spatial scales and conditions. Therefore, the environmental NGRA should be complemented with innovative proposals for addressing the hazard identification requirements.

The new paradigm should start from the regulatory need, supporting the regulatory decisions needed for individual chemicals and proposed uses, but exploring other options than those currently used. For example, the paradigm should support hazard identification needs and the associated risk communication and "downstream" legislative measures but may suggest different hazard classes and categories than those currently used, e.g., informing on the relative vulnerability of taxonomic groups and environmental sub-compartments.

References

Brooks BW, van den Berg S, Dreier DA, et al. (2023) Towards Precision Ecotoxicology: Leveraging Evolutionary Conservation of Pharmaceutical and Personal Care Product Targets to Understand Adverse Outcomes Across Species and Life Stages. Environmental Toxicology and Chemistry 43(3):526-536 doi:10.1002/etc.5754

Chapman PM, Fairbrother A, Brown D (1998) A critical evaluation of safety (uncertainty) factors for ecological risk assessment. Environmental Toxicology and Chemistry 17(1):99-108 doi:10.1002/etc.5620170112

Di Nicola MR, Cattaneo I, Nathanail AV, et al. (2023) The use of new approach methodologies for the environmental risk assessment of food and feed chemicals. Current Opinion in Environmental Science and Health 31 doi:10.1016/j.coesh.2022.100416

Dufourcq Sekatcheff E, Jeong J, Choi J (2024) Bridging the gap between human toxicology and ecotoxicology under one health perspective by a cross-species adverse outcome pathway network for reproductive toxicity. Environmental Toxicology and Chemistry doi:10.1002/etc.5940

Kosnik MB, Schuwirth N, Rico A (2024) Harnessing Computational Methods to Characterize Chemical Impacts on Biodiversity. Environmental Science & Technology Letters 11(3):185-194 doi:10.1021/acs.estlett.3c00865

Newman MC, Ownby DR, Mézin LCA, et al. (2000) Applying species-sensitivity distributions in ecological risk assessment: Assumptions of distribution type and sufficient numbers of species Environmental Toxicology and Chemistry 19(2):508-515 doi:10.1002/etc.5620190233

Rivetti C, Campos B (2023) Establishing a NexGen, mechanism-based environmental risk assessment paradigm shift: Are we ready yet? Integrated Environmental Assessment and Management 19(3):571-573 doi:10.1002/ieam.4763

Rivetti C, Houghton J, Basili D, Hodges G, Campos B (2023) Genes-to-Pathways Species Conservation Analysis: Enabling the Exploration of Conservation of Biological Pathways and Processes Across Species. Environmental Toxicology and Chemistry 42(5):1152-1166 doi:10.1002/etc.5600

Schaupp CM, Maloney EM, Mattingly KZ, Olker JH, Villeneuve DL (2023) Comparison of in silico, in vitro, and *in vivo* toxicity benchmarks suggests a role for ToxCast data in ecological hazard assessment. Toxicological Sciences 195(2):145-154

Schumann P, Rivetti C, Houghton J, Campos B, Hodges G, LaLone C (2024a) Combination of computational new approach methodologies for enhancing evidence of biological pathway conservation across species. Science of The Total Environment 912:168573 doi:https://doi.org/10.1016/j.scitotenv.2023.168573

Schumann PG, Chang DT, Mayasich SA, Vliet SMF, Brown TN, LaLone CA (2024b) Cross-species molecular docking method to support predictions of species susceptibility to chemical effects. Computational Toxicology 30:100319 doi:https://doi.org/10.1016/j.comtox.2024.100319

Smith EP, Cairns J (1993) Extrapolation methods for setting ecological standards for water quality: statistical and ecological concerns. Ecotoxicology 2(3):203-219 doi:10.1007/BF00116425

Tarazona JV (1997) The identification of thresholds of acceptability and danger: the biological route. Arch Toxicol Suppl 19:137-46 doi:10.1007/978-3-642-60682-3_12

2.2.5. Integration of Human Health and Environmental assessment

Strong scientific evidence supports an integrative approach (e.g., One Health) in policymaking and interventions[49], including the increased use of information to evaluate chemical risks to humans and the environment. Health and ecological systems are deeply interconnected, with risks and impacts often overlapping between the two areas through the disruption of similar biological processes. Non-animal approaches offer opportunities to adopt a more unified approach to identifying, predicting, and managing the consequences of chemical exposures by assessing chemical toxicity-based chemical-biology interaction and their multifaceted impacts on biological systems.

The shift towards integrated assessment methods begins with replacing traditional animal tests by leveraging significant technological advancements in bioscience, including molecular toxicology, systems biology, and exposure science. These advancements have resulted in the development of data-rich tools and models that can clarify the key mechanistic pathways linking chemical exposure to adverse health effects in both humans and ecological receptors.

Transitioning to integrated assessment strategies also involves addressing the conceptual and socio-technological obstacles that stem from a conventional dependence on representative animal species, which imperfectly serve as substitutes for humans and natural populations. Greater collaboration between human health and environmental scientists in this process should enhance the overall scientific robustness, reproducibility, and translational relevance of risk assessments, ultimately supporting regulatory frameworks that are adaptable, transparent, and protective of public health and the environment.

In addition to promoting ethical goals such as reducing animal testing, integrated assessments grounded in mechanistic evidence of chemical toxicity reduce duplication, encourage innovation, and support the EU's strategic aims of leading on sustainability and competitiveness.

Considering the many opportunities presented by scientific, regulatory, economic, and ethical imperatives, promoting further integration of human and environmental safety should be a key objective. This goal should, therefore, be fully incorporated into all relevant initiatives arising from this roadmap.

Cross-cutting recommendations, while essential for further integrating Human and Environmental safety assessments, are also relevant to other parts of the roadmap.

- Increase non-animal approaches familiarity through education and training.
- Support the development of comprehensive professional and academic training programmes that integrate human health, environment, and regulation.
- Ensure that the Common Data Platform under one substance, one assessment (OSOA) contains high-quality human and environmental data, including complex datasets such as from non-animal approaches, OMICS, exposure, and others.
- Develop plans for the economic viability of non-animal approaches by creating opportunities for private investors and innovators to compete in providing non-animal approaches services and products that are regulatory fit for purpose and relevant to human and environmental safety.

Background

Rationale for Change – justifying a transition to non-animal approaches that are relevant for the protection of human health and the environment.

[49] European Commission: Group of Chief Scientific Advisors and Directorate-General for Research and Innovation, One Health governance in the European Union, Publications Office of the European Union, 2024, https://data.europa.eu/doi/10.2777/8697309

The European Commission's Group of Chief Scientific Advisors (GCSA), through the Scientific Advice Mechanism (SAM), issued an opinion on One Health governance in the EU, highlighting the interconnectedness of human, animal, and environmental health.

There is strong scientific evidence supporting the application of a One Health approach to policymaking and interventions, including the better use of relevant information to assess the risks of chemicals for protecting both humans and the environment.

Ongoing regulatory changes present a unique opportunity to integrate human health and environmental risk assessments, with the OSOA approach serving as a facilitator for this more comprehensive and efficient framework.

A shift to an integrated approach using non-animal approaches, and where assessments are based on mechanistic understanding, offers the best chance to effectively and scientifically extrapolate toxicological data from one species to another, including humans.

An integrated assessment using non-animal approaches streamlines and improves both safety and hazard assessments across all relevant regulations, resulting in a general reduction in animal testing across regulatory areas, while fulfilling the EU Commission's dual objectives: simplifying regulations and simultaneously maintaining or even improving protection levels.

Integrating the One Health approach into existing EC health priorities, such as the EU Life Sciences Strategy and the EU Biotech Acts, will strengthen the competitiveness of EU SMEs and encourage sustainable innovation throughout the sector. To make One Health and OSOA operational, human and environmental evidence need a common framing. Integration should explicitly link to ecology and biodiversity research programmes (e.g., Biodiversa+, PARC, projects funded under the Horizon Europe programme) and develop frameworks for coordinated assessments and decisions across regulatory regulations.

Critical core requirements for further integration of human and environmental safety assessment.

Supporting greater integration of human and environmental risk assessment requires meeting several key requirements that are vital for the transition to non-animal approaches and their adoption across different regulatory sectors. These considerations should be addressed during the entire implementation of the roadmap, including scientific, regulatory, and educational efforts. Such considerations include, but are not limited to:

- Ensuring that opportunities for further integration of human and environmental safety are woven into all actions related to scientific, educational, economic initiatives, and regulatory processes of the roadmap as core elements.
- Improving the utilisation of exposure information by enhancing the use of human and environmental (bio)monitoring data.
- Supporting a greater understanding of how toxicologically relevant processes essential to human health and environmental safety are evolutionarily conserved across distantly related species, and if and how those translate into possible adverse effects across organisms.
- Developing integrated, tiered safety assessment frameworks that maximise data integration across human and environmental safety, including exposure information.
- Establish cross-sectorial organisational structures to foster closer collaboration among human health and environmental scientists to overcome conceptual, language and cultural barriers to unite these disciplines under the banner of 21st-century toxicology.

- Enhancing confidence in alternatives to animal testing and their applications for human health and the environment by developing case studies and promoting regulatory alignment with the scientific evidence.

Recommendations

Towards a paradigm shift: To fully realise the goal of enabling broader integration of human and environmental safety, a progressive regulatory framework is essential to support a transition to an integrated assessment using alternatives to animal testing, aligned with growing confidence and certainty in the results of cross-species extrapolation, which is based on a mechanistic understanding of physiological and ecological similarities and differences for specific regulatory purposes. For environmental assessments reality-checks that leverage the integration of existing data and knowledge from environmental monitoring and ecological research, is an opportunity to further facilitate a transition. Short-, medium-, and long-term objectives can be achieved through education, training, data and knowledge sharing, and scientific development, as outlined below.

The short- and medium-term recommendations are based on existing regulatory frameworks and initiatives; however, they still need transdisciplinary research and the integration of these infrastructures. Longer-term recommendations will require changes to the current regulatory paradigm and policy.

Short-term recommendations

Develop and enable the necessary structures to facilitate further integration of human and environmental safety.

- Identify and support the networks that can build consensus (e.g., ASPIS, EPAA, EU cross-agency One Health task force, and related structures). Define leadership roles and employ change management practices to enable this integration.
- Foster closer collaboration between human toxicologists and environmental scientists to unify these disciplines while recognising and considering their specificities.
- Define the criteria and indicators most useful for monitoring and evaluating the effectiveness of an integrated approach to assessing chemical risks for regulation.
- Short-term research needs which can be delivered by relevant partners, including research organisations, industry consortia and academia:
    I. Define the “biological domains of applicability” for currently available assays.
    II. Identify data streams common to human and environmental species and focus on how to use them in an integration context.
    III. Develop an ontology of data based on non-animal approaches that complies with regulatory requirements and are relevant to human health and the environment.
    IV. Design and develop case studies that illustrate how the integration of human and environmental health would function in practice.

Medium-term recommendations

Support cross-disciplinary research and knowledge exchange.

- Invest in research:
    - Develop a reference list of mechanisms of toxicity and their roles in identifying adverse effects on different taxonomic groups and ecological functions and establish testing strategies that consider their use for both humans and the environment.
    - Identify current TD/TK approaches common to human and environmental species.
    - Identify mechanisms of toxicity which may lead to adverse effects on different taxonomic groups.

- Increase knowledge of toxicity mechanisms, biological roles and ecological interactions for understanding the potential adverse effects at different levels of biological organisation.
- Research in comparative toxicology and cross-cutting areas (e.g., broaden the number and diversity of representative non-animal-based test systems and PBK models).
- Expand research in comparative toxicology and cross-cutting areas (e.g., broaden the number and diversity of representative non-animal-based test systems and PBK models).

Long-term recommendations

- Phasing out animal testing across regulatory sectors by utilising NAMs data that meets regulatory assessment standards for comprehensive protection of human health and the environment.Facilitate a paradigm shift towards tiered safety assessments enabling complete integration of human and environmental data in regulatory risk assessment, e.g. an overarching next generation risk assessment (NGRA) framework which describes how to seamlessly integrate human health and environment workflows.
- Enable legislative change to adopt an integrative approach, including indicators and monitoring programmes to assess the actual level of protection achieved and implement additional measures as necessary.

2.3. Overarching opportunities across Human Health and Environmental Safety

2.3.1. Replacing *in vivo* with *in silico* methods

Applying assessment elements from the QSAR Assessment Framework (QAF) to support increased regulatory use of QSARs

Rationale and background

Risk assessors may face situations where data availability ranges from scarce to abundant. When toxicological data for substances are lacking, limited physico-chemical or toxicity information can be supplemented with QSAR models to support hazard identification and characterisation efforts. These predictive approaches help bridge data gaps by estimating toxicity or identifying analogous substances with established profiles. Furthermore, for certain toxicological endpoints, the performance and coverage of available models are sufficiently robust to warrant their consideration as the default approach for meeting specific regulatory requirements. Historically, the use of QSARs as a primary source of information regarding hazardous properties was limited. This was mainly because developers provided information on model performance (e.g. sensitivity or specificity for qualitative properties and prediction accuracy for quantitative ones), while definition of applicability domain and model limitations was often missing. Consequently, confidence in QSAR predictions was limited because e.g. the information that a model is performing well for 75% of chemicals from the validation set is not helping the user in concluding if a substance of concern will be included in this subset. As a result, this limitation impeded broader acceptance, as both end users and regulatory evaluators lacked comprehensive methodology to assess the reliability of predictions for compounds of interest.

The recently implemented OECD QSAR Assessment Framework (QAF) has not only updated criteria for model validity but also introduced new measures to evaluate the reliability of individual predictions. These newly introduced principles for predicting reliability and associated Assessment Elements (AEs) provide a comprehensive solution for determining whether a given prediction is appropriate and reliable for regulatory applications. Nonetheless, effectively applying these AEs currently demands significant computational expertise and in-depth understanding of the endpoint for which predictions are generated. Accordingly, it is essential to establish a set of objective criteria which, together with the newly developed AEs for model validity and prediction reliability, will enhance the broader adoption of the QAF by both model developers and end-users.

Proposed approach: A project team is planned to be set up to achieve that through the following actions:

- Identify which AEs would require such criteria.
- For each identified AEs propose suitable criteria, check for every criterion if it can be applicable to all properties or needs to be endpoint-specific and whether that criterion needs to be software implementable. As an example: under prediction reliability, the Performance of the model for similar substances (AE 3.4 in the Prediction and Result Checklists of the OECD QAF) requires a clear definition of 'similar substances,' which can encompass structural, physicochemical, and mechanistic similarities. Moreover, this definition cannot be universal as different toxicological properties might be differently affected by various elements of similarity. Therefore, for this AE, there is a need for software implementable, endpoint-specific similarity measures to assess the adequacy of the closest analogues used to assess performance for substances similar to the target substance.
- Build case studies to examine how newly proposed criteria work in practice.
- In case of positive outcome, amend the OECD QAF and promote the solutions within the modelling community.

Planned output(s)

- QAF AEs assessed and amended by a set of suitable criteria,
- Set of case studies demonstrating usability of QAF and newly developed criteria QAF updated to increase its practical value.

Intended outcome(s)

It is anticipated that this initiative will substantially enhance regulatory adoption of QSARs by increasing awareness of the importance of assessing the reliability of individual predictions and providing effective tools for this purpose. The capability to evaluate the reliability of each prediction is expected to bolster regulatory confidence in computational methods. Furthermore, promoting newly developed criteria among software developers will facilitate straightforward assessment in accordance with QAF, as all essential information required will be automatically generated by the software.

References

Benfenati E, Chaudhry Q, Gini G, Dorne JL. 2019. Integrating in silico models and read-across methods for predicting toxicity of chemicals: A step-wise strategy. Environ Int.131:105060. doi: 10.1016/j.envint.2019.105060.

(Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions - Second edition, OECD Series on Testing and Assessment No. 405, OECD Publishing, Paris.

Ratier A, Lopes C, Multari G, Mazerolles V, Carpentier P, Charles S. 2022.New perspectives on the calculation of bioaccumulation metrics for active substances in living organisms. Integr Environ Assess Manag. 18(1):10-18. doi: 10.1002/ieam.4439.

2.3.2. Developing TK and ADME models for regulatory needs

Development of reliable ADME (absorption, distribution, metabolism and elimination) profiling, PBK (physiologically based kinetic) modelling and (Q)IVIVE tailored to regulatory needs

Rationale and background

Understanding what the body does to the chemical "toxicokinetics" (TK) and what the chemical does to the body "toxicodynamics" (TD) are the fundamental dimensions of toxicology. TK reflects the processes by which a chemical substance enters, moves through, is degraded, and ultimately exits the body, either unchanged or as metabolic byproducts, and is fundamental to chemical hazard assessment. This information is obtained through investigating the absorption, distribution, metabolism, and excretion (ADME) of substances.

Traditionally, animal models have been used in toxicology, where both TK and TD properties influence observable outcomes. However, these studies often do not distinguish the specific contributions of each property or their quantitative relationship. This presents a challenge in developing non-animal approaches and comparing them to data from animal research. For example, low toxicity observed *in vivo* may result from a weak toxicodynamic profile of the parent compound, rapid detoxification, or limited systemic exposure. In contrast, high toxicity may result from bioactivation of the parent compound to a toxic moiety or from bioaccumulation. Considering both TK and TD properties of chemicals is necessary when proposing non-animal approaches to reliably predict hazardous properties for chemicals.

Therefore, understanding of these mechanisms is now integral to non-animal approaches that address complex toxicological endpoints, as reliable results require consideration and computational integration of all relevant elements to achieve regulatory relevance. Consequently, an ADME component forms part of all new non-animal approaches. Additionally, PBK models offer means to predict toxicity for hazard characterisation, and QIVIVE enables conversion of effect doses observed in vitro to external exposures (oral, dermal, or inhalation) for hazard and risk characterisation through reverse dosimetry.

Despite the importance of ADME and TK data, many pieces of legislation (such as REACH) have not explicitly required such data, highlighting the need for standardised methodologies capable of reliably assessing these properties across diverse physico-chemical characteristics. In addition, international standardisation and updated OECD guidance are necessary to expand coverage of non-animal approaches suitable for determining ADME properties.

Proposed approach

This **action** will focus on four tasks:

- Standardising methods for reliable TK parameter determination across chemical classes and harmonising them internationally at the OECD level;
- Developing metabolically competent in vitro models and analytical techniques to accurately identify biotransformation products *in vitro*;
- Developing and integrating models in open access tools to assess bioaccumulation potential in humans and other representative environmental species,
- Developing PBK models and QIVIVE tools that can be parameterised using *in vitro* data and *in silico* results to apply them in NGRA.

Standardisation of Methods: As ADME properties are increasingly recognised as critical components in risk assessment, the routine generation of physico-chemical properties and *in vitro* kinetic parameters is crucial. These include partition coefficients (octanol/water, KoW), intrinsic clearance (Clint), unbound fraction (Fu), and absorption rate. In a number of contexts, isoform-specific metabolism for

phase I and phase II enzymes and transporter data are also significant, particularly given that polymorphic enzymes can result in considerable inter-individual variability in human TK and TD. With these parameters, generic or chemical-specific physiologically based kinetic (PBK) models can be used to predict internal concentrations in body fluids (such as blood or urine) or target organs across species (forward dosimetry). Consequently, it will be necessary to identify appropriate models compatible with these properties, including thorough characterisation of the variability and uncertainty associated with model outcomes and applicability domains. These activities will subsequently inform the development of relevant test guidelines and OECD guidance documents and update of the relevant OECD Harmonised Templates (OHTs) to facilitate international harmonisation and broad regulatory acceptance.

Identification of Biotransformation Products: Metabolic detoxification and bioactivation are key processes that can significantly impact toxicological outcomes. Primary hepatocytes are commonly employed to assess hepatic clearance, but these are unsuitable for incubation periods exceeding four hours. In addition, current analytical methods predominantly track changes in parent compound concentration and often lack robustness in detecting and quantifying relevant metabolites. This task will prioritise the identification of well-characterised, metabolically competent *in vitro* models with sustained stability, followed by the selection of suitable analytical techniques capable of detecting multiple biotransformation products using semi- or non-targeted approaches.

Models for Assessing Bioaccumulation Potential: Accurately determining whether a substance may bioaccumulate within organisms is essential not only for PBT assessment but also for predicting systemic toxicity using non-animal approaches. The toxicity of bioaccumulative compounds is frequently underestimated by most *in vitro* methods; therefore, reliable indicators of bioaccumulation potential are crucial to minimise false negatives. This task will primarily focus on optimising PBK modelling approaches for assessing bioaccumulation while using data generated from *in vitro* models developed under other tasks within this action.

Planned output(s)

- ✓ In the short term: Development of OECD Test Guidelines for the generation of metabolic clearance data in human hepatocytes and protein binding (fraction unbound) to predict human blood and plasma concentrations, along with a defined approach for estimating clearance (half-life and other TK properties such as AUC and Css).
- ✓ In the mid-term, develop additional defined approaches to cover dermal and inhalation routes of exposure and develop OECD TGs as well.
- ✓ Stimulation of efforts aiming at generating TK and ADME data from *in vitro* systems.
- ✓ Preparation of an *in vitro* TK package (annex update, validated test methods and guidance for registrants) to be used as standard information requirement in legislation.
- ✓ Future TK packages should cover relevant species for risk assessment, including human but also consider environmentally relevant species, e.g, terrestrial and aquatic vertebrates, birds and mammals. Initialisation of projects aiming at identification and characterisation of models suitable for the identification of metabolites and bioaccumulation potential.
- ✓ Development of an *in vivo* and *in vitro* ADME/TK database using OECD harmonised templates (e.g. OHT 58 and OHT 201 and any new OHTs for newly validated *in vitro* ADME methods) to support the use of ADME/TK data in NGRA across regulations.

Intended outcome(s)

This action aims to promote the widespread adoption and standardisation of *in vitro* ADME/TK methods across various regulatory frameworks. As this is a critical component for future non-animal approaches, the routine generation of TK/ADME data will facilitate the development of New Approach

Methodologies (NAMs) for assessing complex toxicological endpoints. Furthermore, the data produced through this initiative can support existing regulatory processes, thereby enhancing confidence in non-animal hazard assessment strategies. Finally, the capability to identify relevant biotransformation products will allow for greater confidence in positive and negative findings where necessary.

References

Beaudoin R, Benfenati E, Billat PA, Buratti FM, Chiara Dall'Asta, KeyvinDarney, Gianni Galaverna, Luca Dellafiora L, Pedroni L, Hoogenboom R, Lautz L, Louisse J, Roncaglioni A, Testai E, Tebby C, Thépaut E, Vichi S, Zeman F. 2023. Case studies: Generic kinetic and kinetic-dynamic modelling in human subgroups of the population and animal species to support transparency in food and feed safety. EFSA supporting publication2024:EN-9010. 138 pp. doi:10.2903/sp.efsa.2024.EN-9010.

Bossier, H., J. Chau, C. Ndour, M. Varewyck, T. Verbeke, S. Vergucht A Web-based open source tool for Toxicokinetic and Toxicodynamic modelling EFSA Supporting Publications, 17 (11) (2020), p. 1926E, 10.2903/sp.efsa.2020.EN-1926

Carnesecchi E, Langezaal I, Browne P, Batista-Leite S, Campia I, Coecke S, Dagallier B, Deceuninck P, Dorne JLC, Tarazona JV, Le Goff F, Leinala E, Morath S, Munn S, Richardson J, Paini A, Wittwehr C. 2023. OECD harmonised template 201: Structuring and reporting mechanistic information to foster the integration of new approach methodologies for hazard and risk assessment of chemicals. Regul Toxicol Pharmacol. 142:105426. doi: 10.1016/j.yrtph.2023.105426.

Davidson-Fritz SE, Ring CL, Evans MV, Schacht CM, Chang X, Breen M, Honda GS, Kenyon E, Linakis MW, Meade A, Pearce RG, Sfeir MA, Sluka JP, Devito MJ, Wambaugh JF.2025. Enabling transparent toxicokinetic modeling for public health risk assessment. PLoS One. 20(4):e0321321. doi: 10.1371/journal.pone.0321321. PMID: 40238721; PMCID: PMC12002443.

Dorne JLCM, Cortiñas-Abrahantes J, Spyropoulos F, Darney K, Lautz L, Louisse J, Kass GEN, Carnesecchi E, Liem AKD, Tarazona JV, Billat PA, Beaudoin R, Zeman F, Bodin C, Smith A, Nathanail A, Di Nicola MR, Kleiner J, Terron A, Parra-Morte JM, Verloo D, Robinson T. TKPlate 1.0: An Open-access platform for toxicokinetic and toxicodynamic modelling of chemicals to implement new approach methodologies in chemical risk assessment. EFSA J. 2023 Nov 14;21(11):e211101. doi: 10.2903/j.efsa.2023.e211101.

Bronzwaer S, Devos Y, Dorne JCM, Dujardin B, Dumont AF, Heppner C, Fernández EN, Paraskevopoulos K, das Neves CG. 2025. Advancing EFSA's regulatory science: Updated research and innovation needs. EFSA J. 2025 Apr 16;23(4):e220401. doi: 10.2903/j.efsa.2025.e220401. PMID: 40242780; PMCID: PMC12001418.

Hines DE, Bell S, Chang X, Mansouri K, Allen D, Kleinstreuer N. 2022. Application of an accessible interface for pharmacokinetic modeling and in vitro to *in vivo* extrapolation. Front Pharmacol. https://doi.org/10.3389/fphar.2022.864742

Ratier A, Lopes C, Multari G, Mazerolles V, Carpentier P, Charles S. 2022.New perspectives on the calculation of bioaccumulation metrics for active substances in living organisms. Integr Environ Assess Manag. 18(1):10-18. doi: 10.1002/ieam.4439.

2.3.3. Determining the Point of Departure (PoD) and signatures of toxicity for chemicals through measured omics data

Rationale and background

The hazard assessment of complex systemic toxicity endpoints, such as developmental and reproductive toxicity, currently depends on apical adverse effects observed at the organ or organism level. In contrast, most non-animal approaches yield data primarily at the molecular level, often lacking the means to directly demonstrate toxic effects at the apical level. Hence, a robust understanding of the relationship between molecular-level changes and subsequent adverse outcomes is essential for the regulatory adoption of alternatives to animal testing.

A number of challenges remain unresolved: there is no consensus regarding the criteria to assess biological relevance of models used to generate such data, sampling and sequencing methods are variable, substantial differences exist in data (pre)processing workflows, and multiple strategies are employed to derive points of departure or identify toxicological signatures. While recent efforts by OECD WPHA and WNT have led to progress toward standardising sampling procedures, other aspects still require international consensus and harmonisation. Addressing these challenges may facilitate broader adoption of animal-free approaches and has the potential to improve the efficiency of chemical risk assessments. Omics data can be generated in a high-throughput manner, resulting in extensive datasets that may directly identify modes of toxicological action.

Proposed approach

This initiative aims to support international harmonisation efforts by concentrating on three primary areas:

- Establishing criteria to determine the biological relevance of models used for omics data generation.
- Standardising the omics data processing pipeline.
- Reaching consensus on methodologies for identifying toxicity signatures and deriving points of departure (PoD).

Rather than initiating new activities, this action will identify and engage with relevant ongoing projects to ensure that the focus areas outlined above are adequately addressed. For example, the existing OECD working group responsible for developing guidance on omics-based read-across—an extension of the OECD Omics Reporting Framework (OORF)—is well positioned to address key areas 1 and 2. The main questions to consider regarding biological relevance are: how to demonstrate that key toxicological features of concern are addressed by the proposed biological model (*in vitro* assays), and to what extent this evidence should rely on mechanistic or scientific understanding versus empirical data.

For the omics processing pipeline, the OECD Omics Reporting Framework (ORF) has identified core elements of data processing for both transcriptomics and metabolomics, allowing flexibility in the methods used to address these elements. In this action, appropriate data processing techniques for different scenarios will be determined.

For key area 3, the systemic toxicity working group under the OECD WPHA, comprising leading international experts in alternatives to animal testing, is evaluating various approaches to address systemic toxicity and considering diverse strategies for PoD derivation and the identification of high-concern toxicity indicators. In this area, the most important elements will be to agree the assays composition of the battery to address wide toxicological coverage, specific hazard identification, TK/IVIVe and ADME.

Ongoing and future OECD projects are vital for achieving global harmonisation; however, this action is not confined to OECD initiatives. Other international collaborations, such as APCRA, and major scientific projects under Horizon Europe, PARC, CEFIC LRI, SCCS, and US funding schemes will also be systematically evaluated to identify opportunities for supporting these goals.

Planned output(s)

- To develop case studies with input from JRC, ECHA, and EFSA regarding the use of bioactivity dose-response data for relevant in vitro endpoints and submit them to the OECD IATA programme;
- To finalise the OECD Read Across Application Module under OORF;
- To support the development of guidance documents for applying Points of Departure (PoDs) in next generation risk assessment based on alternatives to animal testing, with the intention to reduce and ultimately phase out animal studies;
- To develop a practical workflow and toolbox to integrate *in vitro* and *in silico* non-animal approaches for deriving PoDs based on such approaches (e.g. by jointly work on regulatory specific adaptations of ASPA workflow).

Intended outcome(s)

This action is intended to support the regulatory application of omics and molecular data by first establishing standardised data (pre)processing workflows and agreeing on a set of techniques to derive PoDs and identify toxicological signatures. Case studies are to be used to illustrate the applicability of these techniques for various regulatory purposes. Guidance documents are also intended to promote consistent approaches for data processing among applicants, registrants, and regulators.

References

Abedini JA, Cook B, Bell SM, Chang XQ, Choksi NY, Daniel AB, Hines DE, Karmaus AL, Mansouri K, McAfee E, Phillips J, Rooney JP, Sprankle CS, Allen D, Casey WM, Kleinstreuer NC. 2021. Application of new approach methodologies: ICE tools to support chemical evaluations. Comput Toxicol. 20:100184. https://doi.org/10.1016/j.comtox.2021.100184

EFSA, Bronzwaer S, Devos Y, Dorne JCM, Dujardin B, Dumont AF, Heppner C, Fernández EN, Paraskevopoulos K, das Neves CG. 2025. Advancing EFSA's regulatory science: Updated research and innovation needs. EFSA J. 2025 Apr 16;23(4):e220401. doi: 10.2903/j.efsa.2025.e220401. PMID: 40242780; PMCID: PMC12001418.

Escher, S. E., Partosch, F., Konzok, S., Jennings, P., Luijten, M., Kienhuis, A., de Leeuw, V., Reuss, R., Lindemann, K., & Bennekou, S. H. Development of a Roadmap for Action on New Approach Methodologies in Risk Assessment. 2022. EFSA Supporting Publications.

Friedman PK, Gagne M, Loo LH, Karamertzanis P, Netzeva T, Sobanski T, Franzosa JA, Richard AM, Lougee RR, Gissi A, Lee JJ, Angrish M, Dorne JL, Foster S, Raffaele K, Bahadori T, Gwinn MR, Lambert J, Whelan M, Rasenberg M, Barton-Maclaren T, Thomas RS. Utility of In Vitro Bioactivity as a Lower Bound Estimate of *In vivo* Adverse Effect Levels and in Risk-Based Prioritization. Toxicol Sci. 173(1):202-225. doi: 10.1093/toxsci/kfz201.

2.3.4. Fostering the use of alternatives to animal testing for nanomaterials and nanoparticles risk assessment

Background

EFSA's Nano Guidance documents (EFSA Scientific Committee, 2021a,b) highlight the relevance in using alternatives to animal testing as part of Integrated Approaches to Testing and Assessment (IATA) for evaluating nano-specific toxicokinetic and toxicodynamic considerations. With the aim to develop further recommendations on the use of alternatives to animal testing for nanomaterials and small/nano particles risk assessment, EFSA has funded experimental projects established through partnership with EU MSs and extra EU organisations, and in collaboration with other EU agencies. Some examples are the EFSA NANOCELLUP Project (GP/EFSA/SCER/2020/04) and EFSA NAMs4NANO Project (GP/EFSA/MESE/2022/01). EFSA's NANOCELLUP Project is an example of a recently finalised case study aimed to design an IATA based on alternatives to animal testing to fill data gaps identified during real risk assessment practice. EFSA's NAMs4NANO Project is still ongoing and aims to develop a "Qualification system for NAMs for EFSA" and, in parallel, a number of proof-of-concept case studies. While the ultimate goal would be to establish a system for broader use in the context of alternatives to animal testing for chemical risk assessment in general, nanomaterials risk assessment has been selected as starting point being an area in which the implementation of novel approaches is promising for several reasons, including the general unavailability of adapted OECD Test Guidelines. The knowledge developed from EFSA projects on the use of alternatives to animal testing for nanomaterials risk assessment will be used in the context of the new self-task mandate from EFSA Scientific Committee to work on the update of 2021 EFSA's Nano Guidances (M-2024-00062).

Ongoing action

EFSA started a multiannual project called NAMs4NANO (2023-2027, 5.3M). The project was granted to a large consortium involving several EU MSs and extra EU organisations and includes a partnership with JRC. The main goal of this project is to promote the implementation of alternatives to animal testing in the risk assessment of nanomaterials, and it is subdivided in three Lots (i.e. Lot 1: Review of NAM-based tools for nano-specific risk assessment & Develop a Qualification System for NAMs in EFSA using the experience from EMA & US FDA; Lot 2: Design and conduct a set of (5) case studies using NAM-based IATA to fill data gaps in nanomaterials risk assessment, covering regulatory areas within the EFSA's remit; Lot 3: Design and conduct a set of (5) case studies to improve NAM-based methodologies for nanospecific assessment). The first two Lots are led by BfR (Germany), and the third Lot by ISS (Italy). The first deliverable of this project is a 'Review of NAM-based tools for nano-specific risk assessment' and 'Proposal for a qualification system for New Approach Methodologies (NAMs) in the food and feed sector: example of implementation for nanomaterial risk assessment', which were published as External Report in the EFSA website. Finalisation of the project is expected by 2027.

Cooperation with ECHA/EMA/SCCS is ongoing on this topic to ensure regulatory relevance of ongoing projects and align on methodologies, when appropriate.

Additional recommendations for research

Two new outsourced activities aimed at promoting the implementation of alternatives for animal testing for nanoparticles risk assessment were recently launched and are expected to start by the end of 2025. The projects are related to 'Developing provisions for guidance on New Approach Methodologies (NAMs) data integration and implementing hypothesis-driven Next Generation Risk Assessment (NGRA), with specific focus on the area of food nanotechnology' and 'Integration of New Approach Methodologies in Chemical Risk Assessment: Case Studies Addressing Nanoplastics in Food (NAMs4NANOPlastics)'.

Lastly, capacity building activities are recommended to support stakeholders in the implementation of methodologies for nanomaterials risk assessment.

3. Recommendations and actions for environmental safety assessments

3.1. Acute fish toxicity

Current data requirements relating to acute toxicity

The OECD TG 203 fish, acute toxicity test (AFT) is the most commonly used standardised animal test method to assess acute fish toxicity (OECD, 2019). Juvenile fish are exposed to a range of concentrations of the test chemical in test solutions with regular observation of mortalities. The TG has been updated in 2019 adding as observations also the visible abnormalities (also known as sublethal clinical signs) related to equilibrium (e.g., loss of balance, floating at the surface or sinking), appearance (e.g., weak or dark pigmentation), ventilatory behaviour (e.g., hyper-, hypo- or irregular ventilation) and swimming behaviour (OECD TG 203, Annex 4). This was added in order to improve the ability to predict chemical toxicity and minimise suffering of animals. Acute fish toxicity is expressed in terms of an external concentration of the test substance in the water that is lethal to 50% of the test organisms (LC50) during the duration of the test; internal concentrations in the fish are uncertain. In addition, no mechanistic information can be obtained from the endpoint of mortality. OECD TG 203 can be conducted with a range of fish species.

Recommendations – acute fish toxicity

Short term:

- **Waiving based on more sensitive taxa:** Include waiving options for fish acute toxicity testing in case of chemical classes for which testing with other aquatic organisms from different taxonomic levels are known to be more sensitive[50].
- **In silico (QSARs):** Use as replacement when demonstrated that chemical is within applicability domain.
- **OECD TG 249:** Use the fish cell line acute toxicity test (RTgill-W1 cell line assay) as default for testing for fish acute toxicity when demonstrated that chemical is within applicability domain.
- **OECD TG 236:** Use the fish embryo acute toxicity (FET) test as refinement when demonstrated that chemical is within applicability domain51.

Medium term:

- Increase the knowledge on **sensitivities of species of different taxonomic groups** and trophic levels with the aim to expand possibilities for waiving fish acute testing.
- ***In silico*** **(QSARs):** Expand applicability domains of existing QSARs
- Explore and expand where possible **chemical applicability domains of OECD TG 249 and OECD TG 236** through mapping and prediction model refinement (data review and discussion; data generation, where required) and document substance (classes) not within the chemical applicability domain.
- **Finalise development of the OECD guidance document on IATA for fish acute toxicity;** Recommend introducing in legislation and guidance an approach that replaces fish acute toxicity testing considering the OECD guidance document on an IATA for fish acute toxicity

[50] The action might also require to further clarify for which chemical classes or for which toxicities informaton from other taxa provide sufficient certainty that protection levels are fulfilled. In this case, the action should be considered a mid-term action.

[51] In the long-term, animal-free approaches not relying on embryos are preferred

and the domain of chemical applicability of alternatives. Information on the specify domain of chemical applicability space could be inlcuded in guidance.

- **Weight of evidence (WoE) and defined approach (DA):**[52]
  - Acute fish toxicity D) for hazard classification: Evaluate with external datasets
  - Perform case studies to gain experience with using different lines of evidence
  - Implement acute fish toxicity DA for hazard classification in regulatory frameworks
  - Expand WoE approaches, also by increasing the understanding of MoAs

Long term:

- ***In silico* methods:** Develop new *in silico*/ QSARs with wider applicability domain
- **OECD TG 249:** Assay being completely animal-derived component free for GLP testing
- **OECD TG 236:** Develop a completely animal-free fish acute toxicity assessment to be able to eliminate this test
- **Weight of evidence:** Advance completely animal-free fish acute toxicity weight-of-evidence assessment

General recommendations

- Increase the understanding and consider inherent variability in acute fish toxicity testing to set data driven expectations for validation of alternatives to animal testing and required precision
- Transparent criteria for applicability, and also documentation of excluded substance with development of guidance necessary re how to handle out of domain substances
- Validate existing mechanistic alternatives to animal testing to increase mode of action (MoA) understanding (e.g. QSARs)
- Develop additional mechanistic alternatives to animal testing to increase MoA understanding

Description of method readiness, applicability domain and validation status

***In silico* (QSAR models)** are already available to predict acute toxicity to fish for some chemical classes. In addition, some QSAR tools include mode of action profilers. The applicability domains of QSAR generally cover organic chemicals with log Kow values[53] up to max. approx. 6 (suggests substance is highly lipophilic). Inorganic chemicals, superhydrophobic, organometallic chemicals, ionic chemicals (though ionic surfactants are covered by ECOSAR) and nanomaterials are generally not within QSAR applicability domains. When QSARs are used, appropriate documentation of evaluation is needed in terms of both, the validity of the model itself (assessed according to OECD guidance document on the validation of QSAR models (OECD 2014)) and the reliability of the outcome (e.g. whether the substance is within the applicability domain of the QSAR).

Within the Partnership for the Assessment of Risks from Chemicals (PARC) various models are being developed for different fish species (publication, in preparation).

***In vitro and assays using non-protected life stages**:* Assays for fish acute toxicity include the OECD TG 249 and the OECD TG 236. Whilst methods using (eleuthero-)embryos are in fact animal methods (such as the OECD TG 236), those are non-protected life stages as defined in the scope of Directive 2010/63/EU before they start to be independently feeding larval forms. The FET and RTgill-W1 tests

[52] This also includes integrated approaches to testing and assessments (IATA) and defined approaches (DA)
[53] Octanol-water partition coefficient, measure between hydrophilicity and lipophilicity of a substance, with log Kow values typically between -3 very hydrophilic and +10 extremely lipophillic

have been validated and applied with different chemicals. Both, the FET and RTgill-W1 tests seem to have similar chemical applicability domains and limitations (e.g. neurotoxicity).

Weight-of-evidence approaches: Tiered testing frameworks and integrated testing strategies (ITS) exist for assessing acute fish toxicity, which aim to maximise use of existing tools and data, and integrate different data streams effectively (ECHA, 2023; OECD, 2012). An IATA is under development, and a score-based DA for acute fish toxicity has been published (Macmillan et al., 2025). A recent analysis of acute fish and daphnid data from the EnviroTox database showed that daphnid data may cover limitations of alternative methods (e.g., neurotoxicity), so that daphnid data could potentially safeguard the use of FET and RTgill-W1 testing for environmental protection, and fish testing would only be required in rare cases (Schür et al., 2025).

Threshold approach: The threshold approach for acute fish toxicity exists as a reduction method (OECD GD 126), whereby fish toxicity is initially evaluated at a single test concentration in an OECD TG 203 limit test, with the concentration set based on algae and daphnid toxicity, with additional fish testing only considered if mortality is observed.

alternative approach	OECD TG 126: Threshold approach for acute fish toxicity
Description incl. endpoint(s)	• Threshold approach is based on observation that fish is not always most sensitive group of aquatic organisms (among algae, invertebrates, and fish) (Rawlings et al., 2019) • Fish toxicity is initially evaluated as a single concentration (limit) OECD TG 203 test, which requires fewer fish compared to a full acute fish toxicity study • This single concentration is the threshold concentration (TC) identified based on lowest EC50 value from reliable algae (OECD TG 201) or acute invertebrate (e.g. OECD TG 202 daphnia) toxicity data

References

Bols, N. C., Barlian, A., Chirino-Trejo, M., Caldwell, S. J., Goegan, P., & Lee, L. E. J. (1994). Development of a cell line from primary cultures of rainbow trout, Oncorhynchus mykiss (Walbaum), gills. Journal of Fish Diseases, 17(6). https://doi.org/10.1111/j.1365-2761.1994.tb00258.x

OECD. (2010). OECD GD 126: Short guidance on the threshold approach for acute fish toxicity.

OECD. (2012). Fish toxicity testing framework, Series on testing and assessment, No 171. https://doi.org/https://doi.org/10.1787/9789264221437-en

OECD. (2013). Test Guideline No. 236: Fish Embryo Acute Toxicity (FET) Test. https://www.oecd.org/en/publications/test-no-236-fish-embryo-acute-toxicity-fet-test_9789264203709-en.html

OECD. (2014). Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. https://doi.org/10.1787/9789264085442-en

OECD. (2019). Test No. 203: Fish, Acute Toxicity Test. http://www.oecd.org/termsandconditions/.

OECD. (2021a). Test Guideline No. 249 : Fish Cell Line Acute Toxicity: The RTgill-W1 cell line assay. http://www.oecd.org/termsandconditions/.

ECHA. (2023a). Chapter R.7b: Endpoint specific guidance Guidance on Information Requirements and Chemical Safety Assessment Chapter R.7b: Endpoint specific guidance (Version 5.0; December 2023). https://doi.org/10.2823/161062

Macmillan, D. S., Ambure, P., Aranda, V., Bayona, Y., Bonderovic, V., Dawick, J., Fabre, N., Fischer, S., Hodges, G., Llobet-Mut, Á., Loisel-Joubert, S., Rivetti, C., Roberts, J., Schirmer, K., Serrano-Candelas, E., Serrano Ramón, B., & Stackhouse, R. A. (2025). Addressing the challenges of acute toxicity hazard classification using a non-animal defined approach. Environmental Toxicology and Chemistry, vgaf046. https://doi.org/10.1093/etojnl/vgaf046

Natsch, A., Laue, H., Haupt, T., von Niederhäusern, V., & Sanders, G. (2018). Accurate prediction of acute fish toxicity of fragrance chemicals with the RTgill-W1 cell assay. Environmental Toxicology and Chemistry, 37(3). https://doi.org/10.1002/etc.4027

Rawlings, J. M., Belanger, S. E., Connors, K. A., & Carr, G. J. (2019). Fish embryo tests and acute fish toxicity tests are interchangeable in the application of the threshold approach. Environmental Toxicology and Chemistry, 38(3). https://doi.org/10.1002/etc.4351

Schug, H., Maner, J., Hülskamp, M., Begnaud, F., Debonneville, C., Berthaud, F., Gimeno, S., & Schirmer, K. (2020). Extending the concept of predicting fish acute toxicity in vitro to the intestinal cell line rtgutgc. Altex, 37(1). https://doi.org/10.14573/altex.1905032

Schür, C., Paparella, M., Faßbender, C., Stoddart, G., Baity Jesi, M., & Schirmer, K. (2025). Daphnids can safeguard the use of alternative bioassays to the acute fish toxicity test: A focus on neurotoxicity. Environmental Toxicology and Chemistry, vgaf014. https://doi.org/10.1093/etojnl/vgaf014

3.2. Aquatic and terrestrial bioaccumulation

Introduction

Bioaccumulation assessments are required for hazard evaluation (e.g., PBT/vPvB assessment, classification, and labelling) as well as exposure modelling and risk assessment in wildlife and humans (ECHA 2023a; ECHA 2023b). These evaluations are mandatory under various EU regulations including REACH, Biocidal Product Regulation (BPR), Classification, Labelling and Packaging (CLP), Veterinary Medicinal Product Regulation (VMPR), Food Additives (FA) and Plant Protection Products (PPP). Bioaccumulation data also provides critical insights into long-term ecotoxicity, as the accumulation of substances in organisms may result in internal concentrations surpassing toxicity thresholds and facilitate their transfer through the food web via biomagnification leading to adverse effects and secondary poisoning.

The shift towards alternatives to animal testing highlights the need for reliable, robust, and biologically relevant alternative strategies for aquatic and terrestrial bioaccumulation assessments considering the substances' physico-chemical properties, proposed use and exposure routes. This document outlines current and future options for assessing bioaccumulation without vertebrate testing, aiming to eliminate *in vivo* testing.

Provisions on the need to perform a bioaccumulation assessment differs slightly depending on the area of legislation. If required, information on aquatic bioaccumulation is usually requested, For aquatic species, a bioconcentration factor (BCF) can be determined using OECD TG 305 or OECD TG 321 HYBIT bioconcentration tests. Also other species are used (e.g. mussels) for bioaccumulation assessment.
Several pieces of legislation include the possibility to waive the bioaccumulation assessment if the substance has certain properties. E.g., under REACH the assessment can be omitted if the substance has a low potential to cross biological membranes and/or an octanol-water partition coefficient (log KOW) ≤ 3. In the case of human medicinal products (European Medicines Agency, 2024), actives with log K_{OW} >4.5 require a full PBT/vPvB assessment.

Some pieces of legislation like REACH or CLP apply a weight-of-evidence determination using expert judgement to conclude whether a substance may be bioaccumulative (B) or very bioaccumulative (vB) (ECHA 2023b; ECHA 2024b). Conclusions can also be drawn from information such as terrestrial bioaccumulation data, biomagnification factors (BMFs), field bioaccumulation assessment factors (BAFs), trophic magnification factor (TMFs), QSAR and mass-balance model predictions, as well as *in vitro* biotransformation rates, *in vitro–in vivo* extrapolation (IVIVE), physicochemical properties, toxicokinetic studies, and biomonitoring data. Methods for collecting, generating, evaluating, and integrating various lines of evidence (*in vivo*, in vitro, in silico bioaccumulation and TK data) to assess the bioaccumulation of organic chemicals in aquatic and air-breathing organisms have been developed and described (Arnot et al., 2022, ECHA 2023a, OECD 2024a)

Legislation does not contain mandatory bioaccumulation data requirements for terrestrial organisms. However, in some legislative areas, such information might still be relevant e.g. for food chain exposure modelling and PBT assessments (ECHA 2023a, ECHA 2023b).
Historically, aquatic bioaccumulation, which may be supported with terrestrial bioaccumulation data (e.g. earthworms), had been assumed to be protective of air breathing organisms. This assumption has come under scrutiny since some chemicals not considered bioaccumulative in the aquatic environment are still able to accumulate in air-breathing species, including birds and mammals (ECHA 2022a, Hoke et al 2015). The need to find improved methods to address this has been highlighted in ECHA's Key Areas for Regulatory Challenge (KARC) report (ECHA 2023c).

Current vertebrate test methods used to meet standard information requirements for bioaccumulation testing

The *in vivo* test OECD TG 305 (Bioaccumulation in Fish: Aqueous and Dietary Exposure) is resource-intensive, animal reliant, time consuming and costly. The guideline allows for two exposure routes: aqueous or dietary. Dietary exposure studies derive a BMF instead of a BCF. Dietary exposure studies should be considered for substances for which it is not possible to maintain and measure aqueous concentrations reliably and/or potential bioaccumulation may be predominantly expected from uptake via feed (e.g. low solubility and high affinity of adsorption to suspended solids or sediments leading to reduced bioavailability in aquatic media) (ECHA 2023a and b). For example, maintaining concentrations of highly hydrophobic chemicals (log K_{OW} > 5 and a water solubility below ~ 0.01-0.1 mg/L) in aquatic media can become technically infeasible. Nevertheless, studies using aqueous exposure are preferred (ECHA 2023b) and regulatory classification criteria are based on aquatic BCFs. In addition, the dietary method faces uncertainties in calculating uptake rates for the purpose of calculating a BCF from the BMF and assumes first-order kinetics, which may not suit all substances, especially those with complex behaviours like ionisable organic compounds (IOCs) or those with rapid metabolism.
Since the adoption of OECD TG 305 in 1996, numerous chemicals have been tested, but substantial variability and uncertainty in *in vivo* BCF data remain, as noted by Arnot and Gobas (2006) and Wassenaar et al. (2020). These uncertainties, particularly with empirical BCF measurements, pose challenges when values approach B or vB thresholds (Camenzuli et al., 2019).

There are no specific test guidelines for bioaccumulation in air breathing organisms. Substances which screen based on physicochemical properties as potentially bioaccumulative in air-breathing organisms (see above) may be further assessed via elimination/metabolism half-lives from QSAR or *in vitro* studies (intermediate assessment) or from existing *in vivo* toxicokinetic (TK) studies, such as from OECD TG 417 (Hofer et al, 2021) (definitive assessment) (ECHA 2023b)
However, in the EU, toxicokinetic and metabolism studies (OECD TG 417) are at present not required for REACH at any tonnage, whereas such studies are required for pesticides. In addition, the OECD TG 417 is not designed for bioaccumulation assessment, and while the standard TK study in rats involves usually a single dose, repeated dose may be needed for assessing bioaccumulation in mammals (Hofer et al. 2021).
Novel modelling and non-animal approaches have been also suggested recently and should be validated in near future (ECHA 2022a, Saunders et al. 2023).

Alternative approaches for bioaccumulation

Under Annex XI of REACH, registrants may adapt information requirements under Annex VII to X by using alternative information including weight-of-evidence, QSAR, in vitro, grouping/read-across and exposure-based waiving.

Screening approaches/in-silico/read-across

The use of log K_{OW} is one of the main criteria used to screen for substances suspected to have a high bioaccumulation potential via partitioning to lipid storage. Under REACH, the standard information requirement for bioaccumulation in aquatic species may be omitted if log K_{OW} is below 3. Screening criteria exist for bioaccumulation in air-breathing mammals are based on log K_{OA} and log K_{OW} (log K_{OW}>2 and log K_{OA} >5; (ECHA 2023b). However, these cut-offs only apply when the potential for bioaccumulation of the substance is solely driven by lipophilicity. Therefore, the use of log K_{OW} as a screening criterion is not considered adequate if the substance is surface active or ionisable at environmental pH (pH 5 – 9) as other mechanisms may drive their bioaccumulation potential. For these substances information on the pKa is needed and apparent partitioning information for the charged form. In these instances, the use of membrane-water partitioning information log K/D_{MW}, besides the log K_{OW}, is considered a more biologically relevant alternative and is gaining increased attention. However, there is no standard guideline available for its empirical determination and *in silico* methods

for log K/D_{MW} are still limited, consisting of ppLFERS (poly-parameter Linear Free Energy Relationships), which have been shown to work well for neutral chemicals, to other more complex computational approaches, which can be expensive and computationally slow.

Whilst the use of physical-chemical partitioning properties provides a simple low-cost approach, none of these parameters take into account biotransformation which may result in an overestimation of the bioaccumulation potential of chemicals which are rapidly biotransformed. It should be noted that no harmonised regulatory cut-offs exist worldwide. Furthermore, specific cut-off values may lead to more or less false negative/positive outcomes (Nendza 2018, Gimeno et al 2024). No regulatory cut-offs have been agreed for log K/D_{MW}.

There are numerous *in silico* models commercially or freely available for predicting BCF with defined guidance (OECD 2024b) to assess their robustness and reliability. These predictions, along with structurally based screening profilers i.e. OECD QSAR Toolbox, are accepted as part of a wider weight of evidence approach. In addition to QSAR models for fish BCFs, mass balance models for simulating BCFs, BAFs, BMFs, HL_T and TMFs have been developed. An overview of the different models and how to apply them as Lines of Evidence (LoEs) is outlined in the OECD IATA on Bioaccumulation (OECD, 2024a). However, these models are generally not considered suitable for polymers, highly hydrophobic chemicals, metals and complex UCVBs and many have a heavy reliance on use of log K_{OW} and do not account for biotransformation.

Several validated QSAR models for predicting whole-body biotransformation rate constants or half-lives have been developed and validated following OECD guidance for QSAR application (OECD 2014). These QSAR predictions can also be used and leveraged to refine BCF estimates and/or serve as a comparator to the OECD TG 319 data as outlined in the OECD IATA on Bioaccumulation (OECD, 2024a). These models include BCFBAF v3.01 module in EPISuite which was developed based on a published database of fish biotransformation rate constants (Arnot and Mackay, 2008; Arnot et al. 2008) and the University of Insubria's QSARINS model (Papa et al., 2014). This information can be combined and compared with other lines of evidence within the Bioaccumulation Assessment Tool (BAT) (Armitage et al. 2021). Several *in silico* models to predict biotransformation and half-lives are easily accessible and comparable via the EAS-E Suite platform (Eas-E-Suite - ARC Arnot Research & Consulting).

Under the Partnership for the Assessment of Risks from Chemicals (PARC), deep learning models for predicting bioconcentration factor (BCF) values are being developed. BCFpro, a model that considers test conditions and species used in addition to chemical structure, has been submitted for publication.
The EFSA TK PLATE model (Dorne, 2023) is an open access platform developed by the European Food Safety Authority (EFSA) to enhance chemical risk assessment through toxicokinetic (TK) and toxicodynamic (TD) modelling. This platform predicts how chemicals are absorbed, distributed, metabolized, and excreted in various species, including humans, animals, and ecological species of relevance. Validation of the model is still on-going with one of the main limitations being the limited availability of TK/TD parameters for many species. Further work (long term) is required to develop the platform to include priorities identified by EFSA, ECHA and PARC including bioaccumulation models.

Read across provides an accepted approach with defined guidance (ECHA 2017a, ECHA 2017b, ECHA 2022b, OECD).

***In vitro* assays/IVIVE (aquatic bioaccumulation)**

Biotransformation is one of the major uncertainties in mass-balance or QSAR models to predict BCFs based on log K_{OW}. Thus, *in vitro* assays have been standardized and validated (Nichols et al. 2018) to determine biotransformation rates which can be incorporated into the models to refine BCF predictions. Currently, these *in vitro* assays (OECD TG 319 A/B; OECD 2018a & b) can be used as an alternative to the current fish *in vivo* study as a part of a weight of evidence approach (ECHA Guidance R7c, 2023).

The biotransformation rate (*in vitro* intrinsic clearance; $CL_{IN\ VITRO,\ INT}$) of a test chemical is determined with either primary hepatocytes (RT-HEP) (OECD TG 319A) or liver S9 sub-cellular fractions (RT-S9) from Rainbow Trout (OECD TG 319B) using a substrate depletion approach. Both *in vitro* methods are highly reliable and reproducible (Nichols et al. 2018). No significant differences in performance have been observed between the two assays using hepatocytes or S9 sub-cellular fractions, thus there is no preference for one *in vitro* system or the other (Fay et al. 2017). Enzymatically inactive controls are included in the test protocol to detect potential abiotic losses due to adsorption or volatility ensuring that observed chemical loss in the active system is due to biotransformation. Whilst these assays have been standardized for Rainbow Trout, they can also be applied to other species and to identify biotransformation products as outlined in the Guidance Document 280 (OECD 2018c).

The measured *in vitro* intrinsic clearance can be extrapolated to a whole-body biotransformation rate constant (k_B) using an appropriate *in vitro-in vivo* extrapolation (IVIVE) model with log K_{OW} as an input. The chemical concentration in fish is then simulated in a one-compartment model by Arnot and Gobas (2003) to predict the BCF (OECD 2018c). IVIVE models have been developed for neutral organic chemicals with log K_{OW} values ranging from 2 to 8 (Nichols et al. 2013). The Nichols et al. model is described in OECD Guidance Document 280 (OECD 2018c). An updated and corrected version has been integrated in the Bioaccumulation Assessment Tool (BAT) (Arnot et al, 2023) and in EAS-E Suite (https://arnotresearch.com/eas-e-suite/). Simple one-compartment models are most commonly used and are based on the assumption that the liver is the major organ for biotransformation (Nichols et al. 2013, Krause and Goss, 2018). However, complex multi-compartment models to generate organ-specific rate constants have also been developed (Krause and Goss, 2020). Generally, a simple one-compartment model is sufficient to predict a BCF unless specific information indicates that first-pass effects or direct blood flow from the gastrointestinal tract (GIT) to the liver is of particular importance. In this instance, it may be more useful to use a multi-compartment model. As a different approach, an empirical model was recently developed for neutral organic chemicals correlating *in vitro* biotransformation rates and *in vivo* BCFs (Laue et al 2023).

The classical IVIVE models cannot directly be applied for BCF prediction of IOCs and surfactants, since log K_{OW} is not considered a reliable parameter for these substances. IOCs display different behaviour compared to neutral organic chemicals such as preferential sorption to membranes (phospholipids) and plasma proteins, the influence of pH on uptake and elimination rates, and the greater potential importance of active transport mechanisms (Armitage et al. 2017). A strong correlation between the log BCF and the membrane lipid−water distribution ratio (D_{MLW}, $R^2 = 0.96$) was found for 10 anionic surfactants and biotransformation was identified as the major mechanism of elimination (Ribbenstedt et al. 2021). In these instances, the BIONIC model, (Armitage et al, 2013 & 2017, CEFIC ECO37 project) provides a more appropriate method for calculating BCF, facilitating the use of log K/D_{MW} and log D_{OW} as partitioning descriptors to overcome the limitations of log K_{OW}. Whilst the BIONIC model provides a more suitable alternative, it is a complex model, and further work is required to improve the transparency of and demonstrate robustness and applicability to a wider range of chemicals.

The incorporation of biotransformation rates for BCF predictions results in a better agreement to *in vivo* values compared to predictions solely based on log K_{OW} (Cowan-Ellsberry et al. 2008, Dyer et al. 2008, Laue et al. 2014, Nichols et al. 2013, Nichols et al. 2018). Whilst these approaches are included in current guidance (ECHA 2023a), regulatory experience with *in vitro* data and IVIVE is still limited therefore the resulting BCF estimate can only be used as part of a WoE approach rather than a direct replacement of a *in vivo* measurement. To date, > 250 unique chemicals have been tested in the TG 319A/B assays including a broad range of chemical classes covering a range of hydrophobicity (log K_{OW} ranging from 0.20 to 8.68). Approximately 1/3 of the tested chemicals are IOCs (Embry et al. MS in preparation), for which the 319 B assay has been shown to work well (Chen et al 2016, Droge et al. 2021).

While the OECD TG 319 assays offer an alternative to *in vivo* assays, there remain certain limitations which include uncertainties regarding their applicability to chemicals for which dietary exposure is

more relevant, such as highly hydrophobic substances, the lack of a clearly defined domain of applicability, and the need for easily accessible, up-to-date extrapolation models for BCF predictions. Furthermore, the applicability domain of alternative approaches (*in silico*, OECD TG 319 A/B, OECD TG 321) for vB classification needs to be compared and defined. Some alternative approaches (e.g. OECD 319 A/B) may not be suitable to provide sufficient confidence to determine whether the BCF is >5000 L/kg (e.g. for potentially very bioaccumulative chemicals with a relatively low log Kow (<ca. 5)). Even if these chemicals are not biotransformed in vitro, log K_{OW} dependent BCF predictions may result in BCFs <5000 L/kg. For these substances further testing in e.g. the HYBIT as a higher tier, may be needed. ECHA's PBT Expert group has set up a working group to make recommendations on use of the OECD TG 319 A/B for bioaccumulation assessment under REACH. Target date for the final report is December 2025.

Since the OECD TG 319 A/B as *in vitro* test is easier to use, needs less resources and is substantially faster compared to the TG 321, the TG 319 A/B is considered the more efficient alternative to animal testing to (de-)prioritize B testing. Therefore, it is recommended as short-term goal to include these assays in chemical legislation to enhance the regulatory acceptance of *in vitro* data to assess bioaccumulation in aquatic species.

There are also a number of additional *in vitro* approaches currently under development which warrant further research. These include:

- Use of S9 fractions from the gut to determine biotransformation for chemicals that undergo substantial extra-hepatic biotransformation, for which current 'liver only' approaches fail to account for, leading to over prediction of bioaccumulation in particular for hydrophobic, poorly water-soluble chemicals for which dietary update is the relevant uptake route (Saunders et al 2020). Organ specific biotransformation was observed in S9 subcellular fractions from liver, intestine, gills and brain for three chemicals in Brown Trout and Rainbow Trout (Franco et al. 2024). If there is significant underestimation of the whole-body biotransformation rate (i.e. resulting in an overestimation of the BCF) extrahepatic biotransformation could optionally be determined to refine BCF predictions using multicompartment models in the long term. This method is still under development and requires further standardisation and validation before use.
- Development of *in vitro* methods to measure very slowly biotransformed chemicals. The relatively short lifespan of the hepatocytes and S9 fractions makes these methods unsuitable for substances with slow biotransformation rates. Improvements were reported for S9 fractions expanding the stability by addition of protease inhibitors (Nichols et al, 2021). However, it should be kept in mind that the IVIVE models have been developed for "well-metabolized" chemicals (Nichols et al. 2013) and slow biotransformation rates have only a minor impact on predicted BCFs. Nevertheless, further work to investigate methods to measure biotransformation of very slowly biotransformed chemicals, e.g. using spheroids, is ongoing.
- Use of immortalised Rainbow trout cell lines from liver, gut and gill to determine biotransformation, used in conjunction with PBTK models to predict bioconcentration in the whole fish (Stadnicka-Michalak et al. 2018). This assay provides a relatively simple high throughput approach (dependant on analytical requirements) which can be adapted to allow for the measurement of depletion kinetics and analysis of biotransformation products. To date only a limited number of chemicals have been tested, and further work is required to investigate and compare the metabolic activity of the immortalised cell lines with those of primary cells. Activities of organ specific biotransformation enzymes were demonstrated recently in different rainbow trout cell lines (Martin et al. 2025). Furthermore, standardization and validation are needed to facilitate use of this method in a longer term as part of a weight of evidence approach with the potential to complement and as a long-term goal to replace the current OECD 319 *in-vitro* assays to not any longer rely on fish as source for hepatocytes or S9 fractions.

***In vitro* assays/IVIVE (bioaccumulation in air-breathers)**

As with aquatic bioaccumulation, the use of *in vitro* biotransformation methods using rat liver S9 sub-cellular fractions and hepatocytes (equivalent to the existing OECD 319A/B for Rainbow Trout) has potential for use in a weight of evidence approach for air-breathing organisms. This method has been developed as part of CEFIC-LRI ECO 41 project for neutral hydrophobics (Lee et al. 2022) and has the advantage of being a simple, high throughput assay with a high degree of reproducibility as an alternative approach to the OECD TG 417 rat *in vivo* study which currently has not been optimized for bioaccumulation assessment (Hofer et al. 2021, Arnot et al. 2022).

The pharmaceutical industry adopted S9 assays for early-stage drug metabolism studies, as a cost-effective, high-throughput alternative to hepatocyte and microsome assays, with the added benefit of covering both phase I and II metabolism. Liver cells ad S9 assays from different mammalian species have been further optimized for automation, reproducibility, and use with human-derived material, improving their relevance for predicting human drug metabolism and inter-individual variability. This knowledge could be utilized for further method development. The clear advantage of these *in vitro* tests is that they could, in principle, cover a range of different cell lines from different species to facilitate the assessment of bioaccumulation potential in a more holistic manner without animal testing. Such methods could also be used to predict interspecies differences in metabolism and toxicity.

The BMF is considered more appropriate for air breathing organisms therefore existing IVIVE models (which are predominately built for extrapolating *in vitro* data to a BCF) are also inappropriate. The use of IVIVE models developed for assessing dietary uptake of pharmaceuticals in rats is well established and with some modifications have the potential to provide a suitable approach. Inter species sensitivity also requires some consideration as *in vitro* biotransformation rates between terrestrial organisms has been demonstrated within pharmacological studies. The method could have merit as part of a weight of evidence screening approach (Lee et al 2017, 2022) for air breathing organisms for predicting BMF but still requires a significant amount of work with regards to validation, understanding interspecies sensitivity and use of IVIVE models.

Research into non animal approaches for bioaccumulation assessments in birds is still very much in the early stages of development but is considered to be a key requirement. In August 2024 the NC3Rs CRACK-IT Challenge "Wings of Change" initiative was launched with the specific focus of advancing alternatives to avian animal testing. Although primarily focused on acute and chronic endpoints, knowledge of biotransformation will be required to develop, use and interpret *in vitro* and *in silico* methods. Avian S9 assays are currently under development at the University of Saskatchewan as part of a HESI sponsored project.

Use of other aquatic species

The current information requirement in REACH is for bioaccumulation in aquatic species (preferably fish) therefore any aquatic organism may in theory be considered as relevant. However, established B and vB criteria for fish may be overconservative for invertebrate and may need further investigations. The use of *Hyallela azteca* (OECD TG 321) in bioaccumulation testing has recently (Oct 2024) been approved as a direct alternative to *in vivo* fish testing by the Member States Committee under REACH (ECHA 2024a). Whilst this provides a regulatory accepted alternative under REACH it has not yet been adopted into other regulations. However, for substances where it is not possible to maintain and measure aqueous concentrations reliably and/or where potential bioaccumulation may be predominantly via food, an OECD TG 305 via dietary exposure is still favoured. This is because the OECD TG 321 only allows to test chemicals via the aqueous route (ECHA 2024a).

The test set-up of the HYBIT assay is very similar to the existing TG 305 bioconcentration test. The organisms (n= ca. 1200 - 1500) are exposed to the test chemical via water either in a semi-static or flow through system followed by a depuration phase of ca. 1 to 4 weeks (or longer) depending on the hydrophobicity of the test chemical. According to the TG 321, only male amphipods should be tested involving manual sexing of the animals. Thus, the HYBIT is resource intensive and expensive with similar costs compared to the TG 305. However, this approach has the advantage of avoiding vertebrate

testing. Whilst the method has undergone validation, this was limited to only three stable substances with relatively low log K_{OW} values chemicals (3.74 and 4.1) and one chemical (hexachlorobenzene (log Kow 5.86) which is clearly bioaccumulative (i.e. no biotransformation). Since the HYBIT is relatively new and there is limited data, the reliability and domain of applicability has to be further explored as a short-term goal. Approximately 30 chemicals have been tested in the HYBIT so far which are either non-bioaccumulative with a relatively low log K_{OW} or (very) bioaccumulative. Since biotransformation is an important elimination mechanism and invertebrates may have limited biotransformation capability compared to fish, there is a need to focus on chemicals which are biotransformed in fish. Thus, further work is needed to widen the range of chemicals tested to build confidence in the method across a wider chemical space. Since the OECD TG 319 A/B as *in vitro* test is easier and substantially faster to use, and needs less resources compared to the TG 321, the TG 319 A/B is considered the more efficient alternative to animal testing to (de-)prioritize B testing in fish which should be considered in a tiered approach (see OECD Bioaccumulation IATA, OECD 2024a).

Bioaccumulation tests using other aquatic invertebrates such as molluscs (ASTM E1022-94) may be relevant as part of a weight of evidence approach (ECHA 2023b) due to their ability to accumulate chemicals primarily through direct water exposure (bioconcentration) and ingestion of suspended particles. However, many molluscs lack metabolic pathways to biotransform certain contaminants, potentially leading to overestimation of bioaccumulation. This highlights the issue that different invertebrate species have varied bioaccumulation capabilities due to differences in uptake routes, metabolism, sensitivity and elimination rates. Thus, results from one species may not always be directly extrapolated to another, necessitating a weight-of-evidence approach to account for interspecies differences.

Efforts are also underway relating to the potential use of Daphnia magna bioaccumulation testing as a screening test (UK Environment Agency, 2025a,b).

Integrated Approaches for Testing and Assessment

The recently published Integrated Approach for Testing and Assessment (IATA) for Bioaccumulation (OECD 2024a) utilizes a Weight of Evidence (WoE) framework to assess discrete organic chemicals in aquatic and terrestrial environments. This approach aligns with OECD guidelines and aims to provide clear guidance for stakeholders in collecting and integrating various Levels of Evidence (LoE) for decision-making.

The IATA combines existing OECD guidelines and Weight of Evidence (WoE) principles to facilitate clear and transparent decision-making in various contexts by both regulators and risk assessors.
It offers examples and guidance for evaluators on how to collect and integrate data from public databases and models for both data-poor and data-rich chemicals. The data evaluation criteria within the IATA are primarily developed from existing OECD TG for each LoE (e.g., OECD TG 319 A/B), providing a systematic approach to address uncertainty within each individual LoE. As each LoE is evaluated, the guidance and methods in this IATA provide a systematic, pragmatic, iterative, and transparent process to determine if there is sufficient confidence in the available data and/or WoE approach for decision-making. Additionally, it includes three case studies to demonstrate its applicability in assessing bioaccumulation.

The Bioaccumulation Assessment Tool (BAT) is one way to operationalize this framework, which is meant to enable users to evaluate data from multiple sources, including *in vitro*, *in silico*, and field data, while addressing data gaps for chemicals with limited information (Arnot et al. 2022). Additionally, the Bioaccumulation Estimation Tool (BET) is available online, allowing users to obtain model predictions based on chemical identifiers. The IATA emphasizes the importance of data reliability and uncertainty in the assessment process, following established OECD recommendations for a transparent and systematic evaluation. Guidance and training for regulators should be promoted soon.

The Bioaccumulation IATA is intended to guide the collection, generation, evaluation, and weighing of various types of bioaccumulation data including physical-chemical, in silico, in vitro and *in vivo* data. Central to this IATA are the elements and principles of WoE such as problem formulation and data relevance and reliability.

The IATA includes different LoEs, including a variety of alternatives to animal testing (previously outlined in this document):

- Phys.-Chem. properties (e.g. log K_{OW} as screening criteria for neutral organic chemicals)
- *In silico* prediction of biotransformation and/ or BCF using QSARs
- OECD TG 319 A/B biotransformation assays in hepatocytes or liver S9 fractions from rainbow trout which can be combined with IVIVE or other models to predict BCFs in fish
- Laboratory bioaccumulation data generated in different species, e.g. invertebrates (e.g. OECD TG 321 *Hyalella azteca* bioconcentration test) and/or OECD TG 305 fish *in vivo* test
- Field data on bioaccumulation and trophic magnification

Conclusions

Bioaccumulation assessments can potentially be fulfilled WITHOUT the use of the OECD 305 *in vivo* assay, primarily based on the availability of various QSARs to predict half-lives and/or BCFs, the OECD TG 319 A/B assays coupled with IVIVE models, and the recently accepted OECD TG 321 (OECD 2024a) *in vivo* invertebrate test (HYBIT). These alternative approaches can also be used to support read across approaches and align with the OECD IATA for Bioaccumulation (OECD, 2024a) which promotes a weight-of-evidence approach for aquatic and terrestrial bioaccumulation assessment. The long-term focus should be on fully *in vitro* methods using that utilise immortalised cell lines from a range of aquatic and terrestrial supported by the integration of in silico predictive and modelling approaches, thus eliminating the use
To enable this transition, it is essential to support the training of stakeholders (regulators, industry) in the evaluation of *in vitro* data, applying toxicokinetic and IVIVE models, and understanding guidelines on applicability and limitations. The onus for this should lie with method developers, early adopters, trade organisations and other academic experts. Existing animal data on toxicokinetics, metabolism, or chronic studies for relevant substances or analogues should also be leveraged to minimise unnecessary animal testing.

Recommendations

Short term

- introduce into legislation and guidance, an approach that replaces the *in vivo* fish bioaccumulation testing considering the OECD guidance document on IATA for bioaccumulation (2024a) and the domain of chemical applicability of alternatives. Available alternative methods are *inter alia* the ones listed below. Domain of chemical applicability space should be explicit in guidance.
    - QSARs
    - OECD 319 A/B with IVIVE
    - OECD TG 321
- Evaluate and promote the applicability of OECD 319 A/B with IVIVE as an efficient and cost-effective approach for decision making on the need for further higher tier bioaccumulation testing e.g. OECD 321.
- Amend Annexes VII to X of Regulation (EC) No 1907/2006 to allow the use of validated OECD methods (e.g., 319A/B, 321) which avoid *in vivo* vertebrate testing. Consider OECD TG 319 A/B as lower tier compared to OECD 321 in accordance with the OECD IATA on bioaccumulation.

- Evaluate and standardize IVIVE models to be used in conjunction with OECD 319 A/B for neutral and ionizable organics across a wider chemical space to improve transparency and regulatory confidence. Models should be easily accessible, clearly defined for regulatory applications and systematically account for potential sources of uncertainty in BCF predictions for a broader chemical space.
- Introduce log K/D_{MW} baseline screening cut-off values for surfactants and IOCs, analogous to the log K_{OW}-based criteria for Tier 1 screening of neutral organic chemicals.
- Compare and evaluate applicability domain of alternatives to animal testing (*in silico*, OECD TG 319 A/B, OECD TG 321) for B and vB classification. Further investigation may be needed to define the domain of applicability and possible limitations of currently available alternatives to animal testing for B and/ or vB classification.
- Superhydrophobic substances (Log K_{OW}> 8): More information needs to be gathered on mechanisms, matrices and parameters enabling assessment of bioaccumulation of super-hydrophobic substances. This will allow development of tools and methods for the bioaccumulation assessment of such substances (KARC, 2025).
- Utilise the results and outcome of an ongoing research project funded by the German Environment Agency (Umweltbundesamt, FKZ 3723 64 405 0), running from April 2024 to March 2027. This project specifically focuses on "Bioaccumulation of superhydrophobic chemicals"

Medium term

- Ensure further confidence in the domain of applicability of the OECD TG 321 by reviewing current data on tested chemicals. If necessary, conduct additional testing across a broader chemical space, especially for chemicals that undergo biotransformation.
- Superhydrophobic substances: Improve bioaccumulation assessment of super-hydrophobic substances which feeds into the identification of substances of very high concern (SVHCs) and for classification of substances as PBT/vPvB (KARC, 2025). Phase out vertebrate testing by excluding current OECD TG 305 data requirements where there is sufficient confidence in using alternatives to animal testing as LoEs and their applicability to B/vB criteria. Existing *in vivo* fish bioaccumulation data may still be used for B/vB assessment and to train models.
- Increase confidence in $logK/D_{MW}$, for screening of surfactants and IOCs, via standardisation and validation of experimental methods including the development of a new OECD Test Guideline.
- Define superhydrophobic chemicals and support development of suitable bioaccumulation assessment methods since they cannot be tested with standard in-vivo or in-vitro procedures due to technical challenges.
- Evaluate the role of extrahepatic biotransformation for a wider set of hydrophobic chemicals where dietary uptake is the relevant uptake route. If there is significant underestimation of the whole-body biotransformation rate (i.e. resulting in an overestimation of the bioaccumulation potential) *in vitro* intrinsic clearance determined in gut S9 fractions (or gill S9 fractions) can optionally be used to refine BCF predictions using multicompartment models.
- Evaluate the metabolic capabilities of invertebrates (*H. azteca*) compared to fish particularly for hydrophobic chemicals. If data demonstrate lower metabolic capabilities for invertebrates, the current ECHA B/vB criteria (developed for fish) may need to be adjusted, where BCFs have been determined in invertebrates.
- Standardize and validate *in vitro* methods for biotransformation in air-breathing organisms using rat hepatocytes and liver S9 fractions including the development of a new OECD Test Guideline.
- Support further development and implementation of IVIVE models) to predict bioaccumulation in air-breathing organisms for regulatory applications.
- Continue to research and develop alternatives to animal testing for avian species including S9 and hepatocyte-based assays (NC3Rs CRACK-IT “Wings of Change”). Evaluate and if required,

build on the tiered approach for bioaccumulation assessment in air-breathers which is already integrated in the OECD IATA on bioaccumulation (e.g. building additional case studies focusing on air-breathers).

Long term

- Further validate and develop relevant TK/TD models, e.g. EFSA TK PLATE model, addressing currently uncovered chemical and biological spaces.
- Continue to research and standardize biotransformation assays in immortalized cell lines to move completely away from the need of animals to isolate S9 fractions or hepatocytes and further develop PBTK models for bioaccumulation prediction based on cell line intrinsic clearance.
- Standardize and validate *in vitro* methods for biotransformation and IVIVE extrapolation models in mammals and birds including the development of a new OECD Test Guideline.

References

Armitage JM, Arnot JA, Wania F, Mackay D. Development and evaluation of a mechanistic bioconcentration model for ionogenic organic chemicals in fish. Environ Toxicol Chem. 2013 Jan;32(1):115-28. doi: 10.1002/etc.2020

Armitage, J. M., R. J. Erickson, T. Luckenbach, C. A. Ng, R. S. Prosser, J. A. Arnot, K. Schirmer, and J. W. Nichols. 2017. 'Assessing the bioaccumulation potential of ionizable organic compounds: Current knowledge and research priorities', Environmental Toxicology and Chemistry, 36: 882-97.

Armitage, J. M., et al., 2021. A critical review and weight of evidence approach for assessing the bioaccumulation of phenanthrene in aquatic environments. Integrated Environmental Assessment and Management. 17, 911-925.

Arnot, J. A., Gobas. F. A. P. C. 2003. 'A generic QSAR for assessing the bioaccumulation potential of organic chemicals in aquatic food webs', QSAR and Combinatorial Science, 22: 337-45.

Arnot, J. A., et al., 2006. Screening Level Risk Assessment Model for Chemical Fate and Effects in the Environment. Environmental Science & Technology. 40, 2316-2323.

Arnot, J. A., et al., 2008. Estimating metabolic biotransformation rates in fish from laboratory data. Environ Toxicol Chem. 27, 341-51.

Arnot, J. A., Mackay, D., 2008. Policies for Chemical Hazard and Risk Priority Setting: Can Persistence, Bioaccumulation, Toxicity, and Quantity Information Be Combined? Environmental Science & Technology. 42, 4648-4654.

Arnot, J.A., Toose, L., Armitage, J.M., Embry, M., Sangion, A., Hughes, L., A weight of evidence approach for bioaccumulation assessment. Integrated Environmental Assessment and Management 2022, 00,00,1-19 https://doi/am-pdf/10.1002/ieam.4583

Arnot, J.A., Toose, L., Armitage, J.M., Embry, M., Sangion, A. and Hughes, L. (2023), A weight of evidence approach for bioaccumulation assessment. Integr Environ Assess Manag, 19: 1235-1253. https://doi.org/10.1002/ieam.4583

Arnot JA, Toose L, Armitage JM, Embry M, Sangion A, Hughes L. 2023. A weight of evidence approach for bioaccumulation assessment. Integrated Environmental Assessment and Management19:1235–1253. DOI: 10.1002/ieam.4583.

Camenzuli, L., et al., 2019. Bioconcentration factors for hydrocarbons and petrochemicals: Understanding processes, uncertainty and predictive model performance. Chemosphere. 226, 472-482.

Chen, Y., Hermens, J.L., Jonker, M.T., Arnot, J.A., Armitage, J.M., Brown, T., Nichols, J.W., Fay, K.A. and Droge, S.T., 2016. Which molecular features affect the intrinsic hepatic clearance rate of ionizable organic chemicals in fish?. Environmental Science & Technology, 50(23), pp.12722-12731.

Cowan-Ellsberry, C. E., Dyer, S.D., Erhardt, S., Bernhard, M.J., Roe, A.L., Dowty, M.E., Weisbrod, A.V. 2008. Approach for extrapolating in vitro metabolism data to refine bioconcentration factor estimates, Chemosphere, 70: 1804-17.

Dorne JLCM, Cortiñas-Abrahantes J, Spyropoulos F, Darney K, Lautz L, Louisse J, Kass GEN, Carnesecchi E, Liem AKD, Tarazona JV, Billat PA, Beaudoin R, Zeman F, Bodin C, Smith A, Nathanail A, Di Nicola MR, Kleiner J, Terron A, Parra-Morte JM, Verloo D, Robinson T. 2023. TKPlate 1.0: An Open-access platform for toxicokinetic and toxicodynamic modelling of chemicals to implement new approach methodologies in chemical risk assessment. EFSA J. 21(11):e211101. doi: 10.2903/j.efsa.2023.e211101.

Droge, S. T. J., Scherpenisse P., Arnot, J.A., Armitage, J.M., McLachlan, M.S., von der Ohe, P.C., Hodges, G (2021). "Screening the baseline fish bioaccumulation factor of various types of surfactant using phospholipid binding data." Environmental Science: Processes and Impacts 23: 1930 - 1948

Droge, S. T. J., Armitage, J. M., Arnot, J. A., Fitzsimmons, P. N, Nichols, J.W. (2021) Biotransformation Potential of Cationic Surfactants in Fish Assessed with Rainbow Trout Liver S9 Fractions', Environ Toxicol Chem, 40: 3123-36.

Dyer, S.D., Bernhard, M.J., Cowan-Ellsberry, C., Perdu-Durand, E., Demmerle, S., Cravedi, J.P. 2008. In vitro biotransformation of surfactants in fish. Part I: Linear alkylbenzene sulfonate (C12-LAS) and alcohol ethoxylate (C13EO8), *Chemosphere*, 72: 850-62.

Exposure And Safety Estimation (EAS-E) Suite, Ver.0.97 updated June 2023. https://arnotresearch.com/eas-e-suite/

ECHA (2017a). Read-Across Assessment Framework (RAAF). Reference: ECHA-17-R-01-EN. Publicly available at https://echa.europa.eu/documents/10162/13628/raaf_en.pdf/614e5d61-891d-4154-8a47-87efebd1851a

ECHA (2017b). Read-Across Assessment Framework (RAAF) - considerations on multi-constituent substances and UVCBs. Reference: ECHA-17-R-04-EN. Publicly available at https://echa.europa.eu/documents/10162/17228/raaf_uvcb_report_en.pdf/3f79684d-07a5-e439-16c3-d2c8da96a316

ECHA (2022a). Bioaccumulation assessment of air-breathing mammals: a discussion paper. https://echa.europa.eu/documents/10162/17228/bioaccumulation_assessment_of_air_breathing_mammals_en.pdf/

ECHA (2022b). Advice on using read-across for UVCB substances. Obligations arising from Commission Regulation 2021/979, amending REACH annexes. Publicly available at https://echa.europa.eu/-/new-advice-for-using-read-across

ECHA 2023a Guidance on Information Requirements and Chemical Safety Assessment Chapter R.7c: Endpoint specific guidance Version 4.0 December 2023

ECHA 2023b Guidance on Information Requirements and Chemical Safety Assessment Chapter R.11: PBT/vPvB assessment Version 4.0 December 2023

ECHA 2023c Key Areas of Regulatory Challenge. Reference: ECHA-23-R-08-E

ECHA 2024a. Minutes of the 87th Meeting of the Member State Committee (MSC-87) MSC/M/87/2024, 10 October 2024

ECHA 2024b. Guidance on the Application of CLP Criteria. Reference ECHA-24-G-05-EN. Part 4: Environmental hazards and Part 5: Additional Hazards. Guidance to Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures. Version 4.0 Nov 2024

ECO37: D-BASS – Developing a Bioaccumulation Assessment Strategy for Surfactants – Cefic-Lri

European Medicines Agency (2024). Guideline on Environmental Risk Assessment of Medicinal Products for Human Use. Revision 1, 2024

Fay, K. A., P. N. Fitzsimmons, A. D. Hoffman, and J. W. Nichols. 2017. 'Comparison of trout hepatocytes and liver S9 fractions as in vitro models for predicting hepatic clearance in fish', *Environmental Toxicology and Chemistry*, 36: 463-71

Franco, M. E., R. Schonenberger, J. Hollender, and K. Schirmer. 2024. 'Organ-specific biotransformation in salmonids: Insight into intrinsic enzyme activity and biotransformation of three micropollutants', *Sci Total Environ*, 925: 171769.

Gimeno, S., Allan, D., Paul, K., Remuzat, P., Collard, M. Are current regulatory log Kow cut-off values fit-for-purpose as a screening tool for bioaccumulation potential in aquatic organisms? Regulatory Toxicology and Pharmacology, Volume 147, 2024, https://doi.org/10.1016/j.yrtph.2023.105556.

Hofer, T. et al. (2021). Analysis of elimination half-lives in MamTKDB 1.0 related to bioaccumulation: Requirement of repeated administration and blood plasma values underrepresent tissues. Environment International, Volume 155, 106592.

Hoke, R., Huggett, D., Brasfield, S., Brown, B., Embry, M., Fairbrother, A., Kivi, M., Leon Paumen, M., Prosser, R., Salvito, D., Scroggins, R.. (2016) Review of laboratory-based terrestrial bioaccumulation assessment approaches for organic chemicals: Current status and future possibilities, Integrated Environmental Assessment and Management, Vol 12 (1), p109–122, https://doi.org/10.1002/ieam.1692

Krause, S., Goss, K.U., 2018. 'In vitro - *in vivo* extrapolation of hepatic metabolism for different scenarios - a toolbox', Chemical Research in Toxicology, 31: 1195-202.

Krause S, Goss KU. (2020) Comparison of a simple and a complex model for BCF prediction using in vitro biotransformation data. Chemosphere. doi: 10.1016/j.chemosphere.2020.127048

Laue, H, Gfeller, H., Jenner, K.J., Nichols, J.W., Kern, S., Natsch, A. 2014. Predicting the bioconcentration of fragrance ingredients by rainbow trout using measured rates of *in vitro* intrinsic clearance', *Environmental Science & Technology*, 48: 9486-95.

Laue, H., et al. (2023), 'Bioconcentration Assessment in Fish Based on In Vitro Intrinsic Clearance: Predictivity of an Empirical Model Compared to In Vitro-*In vivo* Extrapolation Models', Environ Sci Technol, 57 (36), 13325-35.

Lee, Y.-S., Lo, J. C., Otton, S. V., Moore, M. M., Kennedy, C. J., Gobas, F. A. P. C., (2017) In vitro to *in vivo* extrapolation of biotransformation rates for assessing bioaccumulation of hydrophobic organic chemicals in mammals. Environmental Toxicology and Chemistry, 36, (7), 1934-1946.

Lee, Y.-S., Cole, T.R., Jhutty, M.S., Cantu, M.A., Chee, B., Stelmaschuk, S.C., Gobas, F.A.P.C. (2022). Bioaccumulation Screening of Neutral Hydrophobic Organic Chemicals in Air-Breathing Organisms Using In Vitro Rat Liver S9 Biotransformation Assays. Environmental Toxicology and Chemistry. https://doi.org/10.1002/etc.5439

Martin, B.P.M., Franco, M.E. & Schirmer, K. Comparative characterization of organ-specific phase I and II biotransformation enzyme kinetics in salmonid S9 sub-cellular fractions and cell lines. *Cell Biol Toxicol* **41**, 37 (2025). https://doi.org/10.1007/s10565-025-09992-8

Nendza, M., Kühne, R., Lombardo, A., Strempel, S., Schüürmann, G. PBT assessment under REACH: Screening for low aquatic bioaccumulation with QSAR classifications based on physicochemical properties to replace BCF *in vivo* testing on fish. Science of The Total Environment, Volumes 616–617, 2018, Pages 97-106,
https://doi.org/10.1016/j.scitotenv.2017.10.317

Nendza, M., Kosfeld, V., Schlechtriem, C. Consolidated octanol/water partition coefficients: combining multiple estimates from different methods to reduce uncertainties in log KOW. Environ Sci Eur 37, 44 (2025). https://doi.org/10.1186/s12302-025-01072-2

Nichols, J. W., Huggett, D. B., Arnot, J. A., Fitzsimmons, P. N., & Cowan-Ellsberry, C. E. (2013). Toward improved models for predicting bioconcentration of well-metabolized compounds by rainbow trout using measured rates of in vitro intrinsic clearance. Environmental toxicology and chemistry, 32(7), 1611-1622.

Nichols, J., Fay, K., Bernhard, M.J., Bischof, I., Davis, J., Halder, M., Hu, J., Johanning, K., Laue, H., Nabb, D., Schlechtriem, C., Segner, H., Swintek, J., Weeks, J., Embry, M. (2018) Reliability of *in vitro* methods used to measure intrinsic clearance of hydrophobic organic chemicals by fish: Results of an international ring trial, *Toxicol. Sci.* , 164: 563-75.

OECD (2010), Test No. 417: Toxicokinetics, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264070882-en.

OECD (2012), *Test No. 305: Bioaccumulation in Fish: Aqueous and Dietary Exposure*, OECD Guidelines for the Testing of Chemicals, Section 3, OECD Publishing, Paris, https://doi.org/10.1787/9789264185296-en.

OECD (2014), Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, OECD Series on Testing and Assessment, No. 69, OECD Publishing, Paris, https://doi.org/10.1787/9789264085442-en.

OECD (2017), *Guidance on Grouping of Chemicals, Second Edition*, OECD Series on Testing and Assessment, No. 194, OECD Publishing, Paris, https://doi.org/10.1787/9789264274679-en.

OECD (2018), Test No. 319A: Determination of in vitro intrinsic clearance using cryopreserved rainbow trout hepatocytes (RT-HEP), OECD Guidelines for the Testing of Chemicals, Section 3, OECD Publishing, Paris, https://doi.org/10.1787/9789264303218-en.

OECD (2018), Test No. 319B: Determination of in vitro intrinsic clearance using rainbow trout liver S9 sub-cellular fraction (RT-S9), OECD Guidelines for the Testing of Chemicals, Section 3, OECD Publishing, Paris, https://doi.org/10.1787/9789264303232-en.

OECD (2018), Guidance document on the determination of *in vitro* intrinsic clearance using cryopreserved hepatocytes (RT-HEP) or liver s9 sub-cellular fractions (RT-S9) from Rainbow Trout and extrapolation to *in vivo* intrinsic clearance. Series on Testing and Assessment No. 280. https://one.oecd.org/document/ENV/JM/MONO(2018)12/en/pdf

OECD (2024), *Test No. 321: Hyalella Azteca Bioconcentration Test (HYBIT)*, OECD Guidelines for the Testing of Chemicals, Section 3, OECD Publishing, Paris, https://doi.org/10.1787/8ac30c4e-en.

OECD (2024a), *Case study on the Use of Integrated Approaches for Testing and Assessment (IATA) for Bioaccumulation - Ninth Review Cycle (2023)*, OECD Series on Testing and Assessment, No. 404, OECD Publishing, Paris, https://doi.org/10.1787/8bc8ad6f-en.

OECD (2024b), (Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions, Second Edition, OECD Series on Testing and Assessment, No. 405, OECD Publishing, Paris, https://doi.org/10.1787/bbdac345-en.

Papa, E. van der Wal, L., Arnot, J.A., Gramatica, P. (2015) Metabolic biotransformation half-lives in fish: QSAR modeling and consensus analysis. Sci. Tot. Environ., 2014, 470–471, 1040.

Ribbenstedt, A., Armitage, J.M., Günther, F., Arnot, J.A., Droge, S.T.J, and McLachlan, M.S., (2022) *In vivo* Bioconcentration of 10 Anionic Surfactants in Rainbow Trout Explained by In Vitro Data on Partitioning and S9 Clearance. Environmental Science & Technology 56 (10), 6305-6314

DOI: 10.1021/acs.est.1c05543

Saunders, L.J., Wania, F. (2023) Cross-Species Evaluation of Bioaccumulation Thresholds for Air-Breathing Animals. Environmental Science & Technology 2023 57 (29), 10491-10500. DOI: 10.1021/acs.est.2c09115

Stadnicka-Michalak, J., Weiss, F.T., Fischer, M. Tanneberger, K., Schirmer, K. 2018. Biotransformation of benzo[a]pyrene by three rainbow trout (*Onchorhynchus mykiss*) cell lines and extrapolation to derive a fish bioconcentration factor, *Environmental Science & Technology*, 52: 3091-100.

UK Environment Agency, 2025a. Chief Scientist's Group report: The water flea as an alternative to fish for bioaccumulation testing, available at The water flea as an alternative to fish for bioaccumulation testing

UK Environment Agency, 2025b. Tender : Daphnia magna Bioaccumulation Project, available at Daphnia magna Bioaccumulation Project - Find a Tender

Viant M.R., Barnett R.E., Campos B., Colbourne J.K., Barnard M., Biales A.D., Cronin M.T.D., Fay K.A., Koehrn K., McGarry H.F., Sachana M., Hodges G. (2024) Utilizing Omics Data for Chemical Grouping. Environmental Toxicology and Chemistry, 43 (10), pp. 2094 – 2104.

https://doi: 10.1002/etc.5959

Wassenaar, P. N. H., et al., 2020. Variability in fish bioconcentration factors: Influences of study design and consequences for regulation. Chemosphere. 239, 124731

3.3. Fish chronic toxicity

The recommendations provided describe general and effect-specific actions for advancing the weight-of-evidence (WoE) for waiving of *in vivo* chronic fish testing based on different alternatives to animal testing as an integrated approach or group of methods.

The present document is intended to support chronic fish toxicity information requirements under different regulations. The term WoE is therefore considered from a broad perspective without intentionally fully aligning with any specific definition or regulatory requirements.

The main chronic fish toxicity effects covered by this document are: growth, developmental, reproductive, behavioural, and survival effects. Research needs and issues that require a paradigm shift are identified and suggested.

Current animal test methods to meet information requirements for chronic fish toxicity

Summaries of the current *in vivo* methods are set out below, with observed endpoints per study displayed in Table 1.

Fish chronic tests might be used to provide information on different endpoints. Depending on legislation, such tests might be triggered by assessments of chronic fish toxicity, endocrine disruption, or for example in the context of the assessment of persistent, bioaccumulative and toxic properties of substances. Opportunities for reducing the number of animal testing can be harvested if fish chronic tests are designed to provide information relevant for the different assessments (e.g. endocrine disruption, growth, survival etc.), which might require modification of standard test protocols. The description of tests below is indication such opportunities.

OECD TG 210: Fish, Early-life Stage Toxicity Tests (FELS)

Test/exposure duration starts as soon as possible after fertilization (before cleavage of the blastodisc) and ends usually 28 – 60 days post-hatch (depending on species). It covers newly fertilized egg, embryonic development, hatching and early stages of growth.

Reporting of relevant effect concentrations and parameters may depend upon the regulatory framework.

OECD TG 234: Fish Sexual Development Test (FSDT)

The FSDT is essentially an extension of the FELS - exposure is continued until sexual maturation is completed (i.e. usually to around 60 days post fertilization depending on the species), with the aim to cover the sex-labile period in which fish are expected to be most sensitive to the effects of endocrine disrupting chemicals that interfere with sexual development (predominantly estrogen, androgen and steroidogenesis modalities).

Opportunities exist to meet FELS data requirement while investigating EAS-mediated endocrine activity via the OECD TG 234 (Estrogen-Androgen-Steroidogenesis (ECHA and EFSA, 2018)). For this purpose, the OECD TG 234 (point 30) recommends five test concentrations if the data are to be used for risk assessment purposes.

OECD TG 215: Fish Juvenile Growth Test

Some regulatory frameworks (e.g. REACH) may accept the OECD TG 215 as a long-term fish test if there is strong evidence that growth inhibition is the most relevant endpoint in fish (e.g. based on mode of action) and that early-stage exposure is not needed (e.g. if embryonic or larval exposure is not relevant or if these are not the most sensitive stages).

OECD TG 229 (EPA compliant version OPPTS 890.1350) and OECD 230: Fish Short Term Reproductive Assays

These are 21-day group-spawning assays with adult fish. Growth endpoints are therefore not relevant, but mortality can occur due to physiological and behavioural stress arising from reproductive activity in the assay. The tests are usually used as screening assays.

Fish full life cycle test, e.g. OPPTS 850.1500

Test exposure starts with freshly fertilised eggs (before beginning of gastrulation) and covers the hatching of larvae, growth of juvenile fish and reproduction (egg numbers, hatching success and survival of the F1 generation). Depending on the fish species under investigation the study duration is 100-190 days.

It is noted that, the EFSA/ECHA guidance on identification of endocrine disruptors (ECHA and EFSA, 2018) describes that it is possible to modify a fathead minnow Fish full life cycle test study by adapting the test design and including EATS parameters to cover the endpoints required, as recommended in a Medaka Extended One Generation study (OECD TG 240). Particularly the guidance recommends that the assessment of gonad histopathology (e.g. staging of gonads, severity of intersex) is systematically performed each time that the study is carried out. Furthermore, the Zebrafish Extended One Generation Reproduction Test (ZEOGRT) is undergoing validation (Teigeler et al., 2021).

Table 1: Summary table *in vivo* and alternative tests

Observed effects		*In vivo* Methods								Available Alternative Methods					Other approaches			Alternative Methods under development		
		OECD 210 ELS	OECD 234 Sexual development	OECD 229 Short term	OECD 230 Short term	OPPTS 850.1500 FFLC	OPPTS 890.1350 Short term	OECD 240 MEOGRT	OECD 215 Juvenile growth	RaX	TTC	RTgill-W1 cells	OECD 236 FET	QSARs* (e.g. ECOSAR, KATE, VEGA)	ACR	Lack of exposure/ waiving	AOPs	In vitro repro batteries (EAWAG)	Glial cell line	Brain fibroblast-like cell line
Growth	Dry weight	X								X	X	X				X				
	Wet weight	X	X					X	X	X	X	X				X				
	Standard length	X	X					X		X	X	X				X				
	Total length	X						X		X	X	X				X				
Development	Hatching	X	X					X		X	X		X			X				
	Morphological abnormalities Deformed larvae	X	X			X		X	X	X	X		X			X				
	Swim-up	X	X							X	X					X				
Reproduction	Fecundity			X		X		X		X	X					X				
	Fertility					X	X	X		X	X					X		X		
	Sex ratio		X	X		X		X		X	X					X				
Behaviour		X	X	X	X	X	X	X	X	X	X					X				
Survival		X	X	X	X	X	X	X		X	X	X	X			X				
Chronic toxicity (not specific)														X		X				
EDs																	X	X		

*Some QSARs are still under development

Recommendations

Short term:

Several short-term options have been identified that could help to reduce the need for fish chronic toxicity. The Commission intends to clarify the below points with support of EU agencies and discuss them with Member State authorities and stakeholders.

- General (chronic fish toxicity): Review possibilities to waive fish chronic toxicity testing based on mode of action or based on available data from aquatic organisms from different taxonomic groups than fish (e.g. Daphnia).
- Behavioural effects: Clarification on regulatory requirements to report behavioural effects
- General (chronic fish toxicity): Review possibilities to use exposure-based waiving of the chronic fish endpoint, permitted within some regulatory domains, via case studies and additional guidance.

Milestones:

1. A report summarising the discussion with Member State authorities and stakeholders.
2. Update of guidance documents and legislation.

Medium term:

The implementation of the following three recommendations will be explored together with Member States and stakeholders:

- General (chronic fish toxicity): Develop case studies of how different alternative approaches can be strung together (as integrated approaches) to gain better understanding on the:
  - (i) Strengths and limitations of existing methods (in silico, in vitro, and *in vivo*);
  - (ii) Links to other endpoints (acute, non-fish species, endocrine, etc.);
  - (iii) Information that these methods can provide regarding chronic mechanism / modes of action (e.g. by use of TK/TD, QIVIVE).
- General (chronic fish toxicity): Explore possibilities to develop a reference dataset of *in vivo* chronic fish effects / endpoints to:
  - (i) Identify which chronic effects and endpoints are more sensitive per different chemical classes;
  - (ii) Improve the understanding of the inherent variability of *in vivo* chronic fish assays (which will support the validation of alternative approaches);
  - (iii) Generate Acute to Chronic Ratios (ACRs) for different fish species and chemical spaces to better understand where ACRs can be used as parts of WoE-approaches;
  - (iv) Develop integrated testing strategies covering different effects / endpoints.
- General (chronic fish toxicity): Explore the possibilities to develop a reference dataset relating chronic invertebrate (daphnid) effects / endpoints to *in vivo* chronic fish effects / endpoints and to identify which chronic effects and endpoints are more sensitive per different chemical classes across invertebrates and fish;

The following medium-term recommendations are rather research oriented. The Commission will explore possibilities for implementing the recommendations, together with partners from Member States and stakeholders, or e.g. via research funding:

- Growth effects: Investigate further the predictive capacity of fish cell line proliferation for effects on fish growth, by combining in vitro and in silico approaches (TK and growth models). This includes testing further chemicals in the RTgill-W1 cell-line test, but also using other fish cell lines, cell culture models and effect endpoints.
- Growth effects: Support extrapolation of in vitro data to fish *in vivo* effects by comparing short and long-term cytotoxicity and cell proliferation in the RTgill-W1 cell line or on other relevant fish cell lines to chronic fish toxicity data, with focus on survival and fish growth effects.
- Developmental effects: Investigate further the predictive capacity of the OECD TG 236 (FET) for developmental effects by investigating the relationship between additional endpoints in the FET test and chronic fish toxicity, e.g. between teratogenicity endpoints and chronic fish developmental effects
- Mechanistic information – relevant for all chronic effects: Include additional endpoints in the FET test to obtain mechanistic information – e.g. transcriptomics endpoints for tPOD (transcriptomic Point of Departure) derivation, strengthening read-across and derive effect concentrations.
- Behavioural effects: Depending on regulatory requirements, explore the predictive capacity of alternative approaches for chronic fish behavioural effects (e.g. investigate if the OECD TG 236 can be adjusted for incorporating endpoints of relevance for prediction of chronic fish behavioural effects).

Long-term:

The following long-term recommendations are research oriented. The Commission will explore possibilities for implementing the recommendations, together with partners from Member States and stakeholders, or e.g. via research funding:

- Developmental effects: In vitro/ in silico: Develop and validate in vitro and in silico approaches to predict effects on fish development.
- Growth effects: Continue research to increase the functionality of RTgill-W1 assays by advancing experimental procedures as needed, e.g. on adding water flow and cell painting (Fenton et al., 2023; Nyffeler et al., 2025).

Description of method readiness, applicability domain and validation status

Alternative approaches for chronic fish toxicity waiving and weight-of-evidence

This section describes the currently accepted waiving approaches and alternative approaches that can be used to support weight-of-evidence for chronic fish toxicity assessment as integrated approach/ group of methods. Although there is no one-to-one replacement, summary information is provided for the specific available alternative approaches per respective relevant effect/ endpoint.

The information under this section might be regarding also from the perspective of recommendations for reconsidering the data requirements for chronic fish (i.e. by re-examination of the actual need for providing data on these different endpoints).

Since this section intends to reflect on chronic fish toxicity information requirements under different regulations, WoE is therefore considered from a broad perspective without intentionally fully aligning with any specific definition.

Chronic fish data waivers and using other information

Use of available data

Use of available data is promoted or is being considered under different regulations by:

- Encouraging data sharing among applicants. Applicants are encouraged to share environmental risk assessment study data on the same active ingredients. Regarding pharmaceuticals this can be the case between originator and generic companies. Data sharing is specifically mentioned in the revised ERA guideline for human medicinal products (EMEA/CHMP/SWP/4447/00 Rev. 1). The REACH regulation contains obligations for sharing existing data involving vertebrate animals (REACH Art. 27). The advantages are that double testing can be avoided and therefore animal testing can be reduced, and costs can be saved.

- Use of literature data**:** Regulatory assessments, including for fish chronic toxicity, can be based on literature data only, if relevant data are available. For PPPs, Regulation (EC) No 1107/2009 requires dossiers to include a systematic assessment of scientific peer-reviewed open literature on the active substance and its relevant metabolites published within the last 10 years before the date of submission of the dossier. The endpoints taken from publicly accessible publications must fulfil certain criteria and must be assessed for relevance and reliability. This can be done based on standardised assessment methods like e.g. CRED (CRED: Criteria for reporting and evaluating ecotoxicity data (Moermond et al., 2015). The advantages are that no new studies need to be performed, and the amount of costs, labour and animals needed can be reduced. Challenges are to find literature studies which suits the regulatory purpose e.g. provides enough background and raw data and are of high quality. Contradictory results from multiple studies might create interpretation challenges.

- Adaptation of data requirements: REACH does not include an endpoint specific waiver for chronic fish toxicity, but there are general rules for adaptation of standard information requirements (REACH Annex XI). These include for instance the use of existing data, weight of evidence approaches, (Q)SAR, in vitro methods, grouping and read across approaches, or that testing is technically not possible.

Grouping and Read-across

For some modes of action available data indicating that toxicity in chronic toxicity relates to the same mechanism causing acute toxicity (e.g. for Succinate dehydrogenase inhibitors), read-across supporting waiving of chronic based on (slightly prolonged) acute fish tests might be possible.

Such an approach can be supported by validation of TK/TD models (e.g. GUTS) parameterised with acute toxicity data against chronic study data. Toxicity (survival) over long-term and time-varying exposure can then be modelled, avoiding the use of chronic studies for substances with the same mode of action. This approach may also be supported by modification of acute tests to better inform time-to-effect parameters for subsequent TK/TD modelling, at least for some known modes of action.

Exposure based / environmental fate-based waiving

Under Regulation (EC) 1107/2009, chronic fish toxicity is required for active substances where exposure in edge-of-field surface waters is likely, and the substance is stable in water (<90% loss by hydrolysis in 24 h).

Under Regulation (EU) 528/2012, chronic fish toxicity testing is required if the results of the ecotoxicological studies, studies on fate and behaviour and/or the intended use(s) of the active substance indicate a risk for the aquatic environment, or if long-term exposure is expected.

Under REACH, the general rules for adaptation in Annex XI allow for substance-tailored exposure-driven testing (Section 3 of Annex XI). Testing in accordance with Annex IX (the Annex at which chronic fish toxicity testing is required) may be omitted based on the exposure scenario(s) developed in the Chemical Safety Report, provided any of the criteria in Section 3.2 of Annex XI are fulfilled (i.e. there is an absence of or no significant exposure, a PNEC can be derived from the available test data, and exposures are always well below the derived PNEC; or strictly controlled conditions are applied throughout the life cycle) and adequate justification and documentation is provided, and the specific conditions of use are communicated through the supply chain.

Waiving of chronic fish based on mode of action

The identification of the most sensitive organism / taxonomic group based on mode of action, and accordingly waiving of chronic fish testing when fish are not the most sensitive organism, is recommended. For example, the need and waiving of chronic fish testing may be relevant on a mode of action basis for certain pesticides when the risk is driven by other organisms due to mode of action e.g. herbicides and insecticides. Initiatives to deprioritise and waive chronic fish tests include:

- Waiving of chronic fish testing in case fish are not the most sensitive species:
    - The US EPA Office of Pesticides Programs (OPP) allows waiving chronic fish toxicity testing of pesticides if there is no acute toxicity and little exposure, unless there is high bioaccumulation.
    - As reported by Teixido et al. (2023), the FELS toxicity (OECD TG 210) endpoints were the most sensitive (10-fold) for 9.5% of analysed chemicals (n=223) when compared to chronic Daphnia and algae endpoints. Some of these compounds have a known or assumed specific mode of action such as endocrine disrupting, or other non-narcotic MoA, indicating that the higher sensitivity in the FELS toxicity test is related to a specific mechanism of action. Thus, the use of existing data (i.e., chronic non-vertebrate data, and data from alternatives to animal testing) may improve (and reduce the uncertainty associated with) assessments for a more integrated approach to evaluate chronic fish toxicity.
    - For neurotoxic chemicals, Schür, Paparella et al. (2025) analysed historic acute toxicity data for fish and daphnids and concluded that fish rarely represent the most sensitive trophic level of the two, a conclusion which could be explored for considering also chronic fish toxicity.

    - For human pharmaceuticals with a specific toxicity profile (e.g. endocrine active substances, antibacterials) a so-called tailored testing strategy is applied, which differs from the standard test composition. In the revised ERA guideline for human medicinal products (EMEA/CHMP/SWP/4447/00 Rev. 1) fish testing (e.g. OECD 210) for antibacterials.
- Decision tree for reducing fish testing for legacy pharmaceuticals: A decision tree has been proposed by Coors et al (2023) to reduce the need for experimental testing on fish for legacy pharmaceuticals. If substantial evidence indicates that a pharmaceutical may affect fish, *in vivo* testing is warranted. In cases where such evidence is lacking, the decision tree employs a risk-based approach using two proxies to assess hazard, and both are compared to the predicted environmental concentration (PEC) as a conservative estimate of environmental exposure. The first proxy for characterising hazard is the minimum no observed effect concentration (NOECmin) derived from non-vertebrate chronic toxicity tests (OECD TG 211 and OECD TG 201). The second proxy is derived by the fish plasma model (FPM) that suggests toxic effects in fish are unlikely below human therapeutic plasma concentrations. In the next step, both of these proxies for hazard (NOECmin and TWC, theoretical therapeutic water concentration) are related to the chemicals' PEC in surface water. The two derived risk quotients (PEC/NOECmin and PEC/TWC) are then compared to a proposed threshold value of 0.001. If at least one of the two risk quotients is greater than the threshold value, the compound should be considered for *in vivo* testing in fish. If both quotients are below the threshold value, no *in vivo* fish testing has to be considered. The threshold value of 0.001 which correspond to a safety factor of 1000 may be adjusted in the future based on new data.

Extrapolation approaches

Acute to Chronic ratios

Acute-to-chronic ratios (ACRs) can be used to estimate chronic toxicity of chemicals in cases where only acute toxicity data are known, and chronic data are either limited or absent (Raimondo et al., 2007). They are empirically derived as ratios between the acute (represented by LC50 or EC50) and chronic (represented by NOEC or MATC) ecotoxicological thresholds (Brill et al., 2021; Raimondo et al., 2007).

Extrapolations for fish chronic endpoints that are based on acute fish or Daphnia toxicity have been shown as a valid method by comprehensive data analysis provided by the EU Commission Joint Research Centre (Kienzler et al., 2016). In general, the use of acute to chronic ratios has been demonstrated as a valid replacement in the scientific literature (e.g. Brill et al., 2021; Raimondo et al., 2007, Scholz et al., 2018, Wheeler et al., 2014), particularly is the mode of action of chemicals has been accounted for. Whilst values can vary, if the MoA of a chemical is narcotic, its ACR is usually relatively constant (Wang et al., 2020).

If reliably defined, ACRs can be used as screening tools or weight of evidence to waive need for chronic testing. However, one of the key challenges is the fact that robust and reliable data is often lacking, particularly with respect to existing fish chronic toxicity data. Ideally, data should be matched by species to avoid issues around species sensitivity differences. No single ACR exists for all chemicals, leading to a lack of standardisation and potential challenges in knowing which factor to use.

Currently, ACRs are not accepted by ECHA under REACH as a direct replacement for *in vivo* studies. Use of ACRs is most appropriate for chemicals with a narcotic mode of action under both acute and chronic exposures, thus their use also requires reliable determination of MoA.

Given the inherent challenges of deriving robust ACR (data availability, a priori knowledge of MoA) it is unlikely that this approach will ever be considered reliable enough to be considered a direct replacement for an *in vivo* fish chronic study. However, the approach could have some merit as a screening tool or as part of a wider WoE argument. It is recommended to generate and publish more ACRs for different fish

species and chemical spaces to better understand where ACRs can be used as parts of WoE approaches including under REACH.

Alternative approaches to animal testing for assessment of chronic fish toxicity including standardised stage and applicability domain

Several alternative approaches to potentially predict chronic effects in fish are already available. However, they are in earlier stages of development than acute alternatives and require additional development and data collection efforts.

This section summarises the information on available alternative approaches according to respective relevant effect / endpoint, including also approaches that are still in an exploratory phase. It is expected that effect-specific WoE can be developed for some / each of the main effects / endpoints considering different alternative approaches (integrated approach / group of methods).

Recommendation: As regards fish chronic endpoints, there is a strong need to develop case studies of how different methods can be strung together to gain better understanding on the strengths and limitations of the existing methods (in silico, in vitro, and *in vivo*), links to other endpoints (acute, non-fish species, endocrine, etc.) and information that these methods can provide regarding chronic mechanism / modes of action. A better understanding of different methods working together would also be required for the development of an IATA or Defined Approach.

Growth

Stadnicka-Michalak et al. (2015) presented an approach based on a fish cell line assay and PBTK modelling that quantitatively predicts fish growth effects. The authors exposed RTgill-W1 cells for 120 h to two fungicides, measured cell survival and proliferation, predicted long-term inhibition of cell proliferation, interpolated fish weight reduction to different chemical concentrations, and presented model predictions of reduced fish growth in agreement with *in vivo* data. The model is so far only applicable to rainbow trout and fathead minnow. The approach would need to be further developed before use for regulatory assessments is possible. In particular, data with further substances are required.

Survival

Hernández-Moreno et al. exposed RTgill-W1 cells to 15 nanobiomaterials (NBM) and measured cytotoxicity (viability) for up to 28 days. The cells can be used for screening of long-term toxicity of nanobiomaterials. The approach evaluates the time needed to reduce by 20% or 50% the cell viability with respect to a control. To consider applying the approach in regulatory assessments, a comparison with the long-term survival of fish *in vivo* would be necessary.

Development

In addition to acute mortality, the OECD TG 236 allows for observation of hatching and lethal as well as sublethal developmental effects in zebrafish embryos up to 96 hours post fertilisation, or possibly longer. To use such data, a recommendation is to investigate the relationship between teratogenicity information from the OECD TG 236 and from chronic developmental fish toxicity.

Another recommendation is to include transcriptomics endpoints in the OECD TG 236 to also obtain mechanistic information for tPOD derivation, strengthening read-across and derive effect concentrations. ECHA contracted the Fraunhofer Institute to explore including transcriptomics endpoints in TG 236 to also obtain mechanistic information for tPOD derivation, i.e. alterations of gene expression, which could strengthen category building for read-across. The aim is to adapt the OECD TG 236 to use it as an

alternative to chronic fish toxicity testing. According to ECHA, the approach could not only be used to derive effect doses from molecular data in fish, but potentially also inform about the mode of toxicological action. It is expected that it could potentially also be used to address endocrine disruption or other concerns which are based on known mechanisms of toxicity.

Reproduction

Reproductive adversity as it is conceived now is not assessable by methodologies other than *in vivo* methods. This is also the case for generating adversity related to endocrine disruption. A better knowledge of adverse outcome pathways (AOP) could help to link information from in vitro methods to adverse effects.

Behaviour

It is recommended, in the short-term, to clarify the regulatory requirements to report behavioural effects. Based on the actual requirements, the predictive capacity of alternative approaches for chronic behavioural effects in fish can be explored in the medium-term by, e.g., investigating if endpoints can be included in the OECD TG 236 for prediction of chronic fish behavioural effects. Furthermore, available literature data could be assessed to explore the performance of fish behavioural studies to be used not only as a screening for neurotoxicity but also for endpoint derivation.

The following considerations are based on that the roadmap aims to phase out animal tests gradually, starting from protected fish life stages[54] of vertebrates (i.e. from fish larvae with external feeding), then later progressing to non-protected life stages (embryos and early non-feeding larvae), and in the future also phasing out testing with invertebrates.

In vitro and eleutheroembryos neurotoxicity assays available

Recent publications have described the development and the application of brain cell lines of marine teleost fish to investigate metal neurotoxicity, namely glial cell line derived from the Red Sea Bream (Luo et al., 2023) and brain fibroblast-like cell line derived from the European sea bass (Morcillo et al., 2017).

AOPs

The opportunities and challenges for defining AOPs that might be linked to neurotoxic and behavioural effects have been explored by review papers (Fitzgerald et al., 2021; Legardi et al., 2018) with a focus on the use of behavioural assays with fish early life stages. Identified research needs to include the identification of molecular targets, validation of pathways, and determination of the molecular events causing the observed phenotype (Fitzgerald et al., 2021). Ultimately, the use of AOPs, also for cross-species extrapolations and ecological risk assessment, could be promoted if behavioural changes were incorporated in AOPs as whole organism responses (Legardi et al., 2018).

In silico

Quantitative structure-activity relationship models (QSAR) are available that predict chronic toxicity to fish, including the OECD QSAR toolbox, ECOSAR, KATE and VEGA. Some QSAR tools like iSafeRat include mode or mechanism of action profilers.

The applicability domains of QSARs generally cover organic chemicals up to log Kow of 8. Inorganic chemicals, organometallic chemicals, ionic chemicals (though ionic surfactants are covered by ECOSAR) and nanomaterials are generally not within QSAR applicability domains.

[54] According to Directive (EC) 2010/63

References

Brill, J. L., Belanger, S. E., Barron, M. G., Beasley, A., Connors, K. A., Embry, M., & Carr, G. J. (2021). Derivation of algal acute to chronic ratios for use in chemical toxicity extrapolations. Chemosphere, 263, 127804.

Coors A, Brown AR, Maynard SK, Nimrod Perkins A, Owen S, Tyler CR. 2023. Minimizing Experimental Testing on Fish for Legacy Pharmaceuticals. Environ Sci Technol. 57(4):1721-1730. eng.

ECHA and EFSA. 2018. ECHA (European Chemicals Agency) and EFSA (European Food Safety Authority) with the technical support of the Joint Research Centre (JRC), Andersson N, Arena M, Auteri D,Barmaz S, Grignard E, Kienzler A, Lepper P, Lostia AM, Munn S, Parra Morte JM, Pellizzato F, Tarazona J, Terron A and Van der Linden S, 2018. Guidance for the identification of endocrine disruptors in thecontext of Regulations (EU) No 528/2012 and (EC) No 1107/2009. EFSA Journal 2018;16(6):5311,135 pp. https://doi.org/10.2903/j.efsa.2018.5311. ECHA-18-G-01-EN.

Fenton PC, Turner CJ, Hogstrand C, Bury NR. Fluid shear stress affects the metabolic and toxicological response of the rainbow trout gill cell line RTgill-W1. Toxicol In Vitro. 2023 Aug;90:105590. doi: 10.1016/j.tiv.2023.105590. Epub 2023 Mar 28. PMID: 36997009.

Fitzgerald JA, Könemann S, Krümpelmann L, Županič A, Vom Berg C. Approaches to Test the Neurotoxicity of Environmental Contaminants in the Zebrafish Model: From Behavior to Molecular Mechanisms. Environ Toxicol Chem. 2021 Apr;40(4):989-1006. doi: 10.1002/etc.4951. Epub 2021 Feb 1. PMID: 33270929.

Hernández-Moreno D, Navas JM, Fernández-Cruz ML. 2022. Short and long-term effects of nanobiomaterials in fish cell lines. Applicability of RTgill-W1. Chemosphere. 309:136636.

Kienzler, A., Halder, M., & Worth, A. (2016). Waiving chronic fish tests: possible use of acute-to-chronic relationships and interspecies correlations. Toxicological & Environmental Chemistry, 99(7−8), 1129−1151. https://doi.org/10.1080/02772248.2016.1246663

Legradi, J.B., Di Paolo, C., Kraak, M.H.S. et al. An ecotoxicological view on neurotoxicity assessment. Environ Sci Eur 30, 46 (2018). https://doi.org/10.1186/s12302-018-0173-x

Luo, X.; Fu, X.; Zhang, M.; Liang, H.; Niu, Y.; Lin, Q.; Ma, B.; Liu, L.; Li, N. Development of a New Marine Fish Continuous Cell Line Derived from Brain of Red Sea Bream (Pagrosomus major) and Its Application to Fish Virology and Heavy Metal Toxicology. Animals 2023, 13, 3524. https://doi.org/10.3390/ani13223524

Moermond CTA, Kase R, Korkaric M, Ågerstrand M. 2015. CRED: Criteria for reporting and evaluating ecotoxicity data. Environmental Toxicology and Chemistry. 35(5):1297-1309.

Morcillo P, Chaves-Pozo E, Meseguer J, Esteban MÁ, Cuesta A. Establishment of a new teleost brain cell line (DLB-1) from the European sea bass and its use to study metal toxicology. Toxicol In Vitro. 2017 Feb;38:91-100. doi: 10.1016/j.tiv.2016.10.005. Epub 2016 Oct 13. PMID: 27746373.

Nyffeler J, Harris FR, Willis C, Byrd G, Blackwell B, Escher BI, Kasparek A, Nichols J, Haselman JT, Patlewicz G, Villeneuve DL, Harrill JA. A combination of high-throughput in vitro and in silico new approach methods for ecotoxicology hazard assessment for fish. Environ Toxicol Chem. 2025 Jan 6:vgae083. doi: 10.1093/etojnl/vgae083. Epub ahead of print. PMID: 39937625.

Raimondo S, Montague BJ, Barron MG. Determinants of variability in acute to chronic toxicity ratios for aquatic invertebrates and fish. Environ Toxicol Chem. 2007 Sep;26(9):2019-23. doi: 10.1897/07-069R.1. PMID: 17705662.

Scholz et al., 2018. Meta-analysis of fish early life stage tests—Association of toxic ratios and acute-to-chronic ratios with modes of action. Environmental Toxicology and Chemistry, Volume 37, Issue 4, 1 April 2018, Pages 955–969, https://doi.org/10.1002/etc.4090.

Schür C, Paparella M, Faßbender C, Stoddart G, Baity Jesi M, Schirmer K. Daphnids can safeguard the use of alternative bioassays to the acute fish toxicity test: A focus on neurotoxicity. Environ Toxicol Chem. 2025 Jan 21:vgaf014. doi: 10.1093/etojnl/vgaf014. Epub ahead of print. PMID: 39836637.

Stadnicka-Michalak J, Schirmer K, Ashauer R. 2015. Toxicology across scales: Cell population growth in vitro predicts reduced fish growth. Sci Adv. 1(7):e1500302.

Teigeler M., Eilebrecht E., Schäfers C., Maack G., Walter-Rohde S. (2021) Population relevant effects on fish – an update on ZEOGRT validation. In Proceedings, SETAC Europe 29th Annual Meeting (p. 10), Virtual meeting, May 3-6, 2021.

Teixido E, Leuthold D, de Croze N, Leonard M, Scholz S. 2020. Comparative Assessment of the Sensitivity of Fish Early-Life Stage, Daphnia, and Algae Tests to the Chronic Ecotoxicity of Xenobiotics: Perspectives for Alternatives to Animal Testing. Environmental Toxicology and Chemistry 39 (1):30-41. doi:10.1002/etc.4607

Wang Z, Berninger JP, You J, Brooks BW. One uncertainty factor does not fit all: Identifying mode of action and species specific acute to chronic ratios for aquatic life. Environ Pollut. 2020 Jul;262:114262. doi: 10.1016/j.envpol.2020.114262. Epub 2020 Feb 24. PMID: 32120260.

Wheeler JR, Maynard SK, Crane M. An evaluation of fish early life stage tests for predicting reproductive and longer-term toxicity from plant protection product active substances. Environ Toxicol Chem. 2014 Aug;33(8):1874-8. doi: 10.1002/etc.2630. Epub 2014 Jun 27. PMID: 24799351.

3.4. Endocrine disruption for environmental safety assessments

Introduction

Endocrine disruptors (EDs) are "exogenous substances or mixtures that alter the functions of the endocrine system and consequently cause adverse effects in intact organisms, or its progeny or population" (WHO, 2002). According to the ED identification criteria as defined in the Guidance to Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures (ECHA, 2024), in the Plant Protection Products (PPP) regulation 605/2018 and Biocical Product Regulation 2100/2017, a substance shall be considered as having ED properties if it meets all the following criteria: a) it shows an adverse effect in an intact organism or its progeny, which is a change in the morphology, physiology, growth, development, reproduction or life span of an organism, system or population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress or an increase in susceptibility to other influences, b) it has an endocrine activity, i.e. it alters the function(s) of the endocrine system, c) the adverse effect is a consequence of the endocrine activity. According to the above-mentioned regulations, a substance is considered to be an endocrine disruptor if the three criteria are met, unless the adverse effects identified are not relevant at the level of population. The CLP guidance (ECHA, 2024) reports that effects on growth (body weight and length), development and reproduction (such as fecundity, fertility, sex ratio, and hatching success) in single species in laboratory studies are generally regarded as adverse at the level of population. Therefore, establishing causal links between molecular-and individual-level responses is crucial to conclude on ED properties as well as to further elucidate extrapolation from individual level effect to population. It is important to recognise that the domain of applicability of the term 'endocrine disruptor' is contingent on the defined scope of the endocrine system. To date, most scientific and regulatory efforts related to EDs have been focused on interference with the estrogen (E), androgen (A), thyroid (T) and steroidogenesis (S) pathways, referred to as EATS modalities. However, the endocrine system comprises other hormone signalling systems beyond EATS modalities, that potentially could also be affected by exposure to xenobiotics – collectively these are considered 'non-EATS modalities'. Potential non-EATS modalities include progesterone signalling, and the Hypothalamic-Pituitary-Adrenal (HPA) and somatotropic axes. The disruption or perturbation of those pathways can potentially lead to a high variety of apical effects involving growth, development, reproduction, metabolism, behaviour, etc. Moreover, under non-EATS-modalities, endocrine pathways relevant for invertebrates, e.g., ecdysone and retinoic acid should be also acknowledged.

The above-mentioned identification criterion (a) includes the notion of 'intact organisms', which is understood to mean that the effect would occur *in vivo*, observable in a test animal system. However, it does not necessarily mean that the adverse effect has to be demonstrated in an intact test animal (ECHA, 2024) but may be shown in adequately validated alternative test systems predictive of adverse effects in humans and/or wildlife (Solecki et al., 2017). Operational actions towards phasing out animal testing for the environmental safety assessment of endocrine disruptors were identified and outlined in the following paragraphs in a stepwise progression. Each step (short-term, medium-term, and long-term) is building upon the previous one to ensure a cohesive and effective approach to reach the overall objective. The recommendations provided apply across all chemical sectors. Nevertheless, the relevance and feasibility of these recommendations may differ between sectors, given the distinct data requirements imposed by different legislative frameworks governing each sector.

Recommendations

Reduction of animal testing

The following options for reduction or refinement of animal testing in the short- to medium term were identified. Reduction of animal testing would be an interim solution until a full transition to an environmental safety assessment that relies fully on the use on non-animal approach is possible.

Adapt test guidelines or guidance so that information from *in vivo* testing is maximised and can be used both for concluding on the ED properties and risk assessment. It is recognised that the testing strategy recommended for the investigation of the endocrine disrupting properties is not fully in line with the testing required for performing a risk assessment. However, the two strategies could be aligned since this will considerably reduce the use of animals. This would require adapting current *in vivo* testing inter alia as regards number of dose levels applied, since the derivation of reliable no observed effect concentrations for use in risk assessment require at least four test concentrations or considered endpoints. Legislative requirements or guidance would need to be modified.

A similar measure would be to **complement the available study protocols with additional endpoints** that can give indications on toxicity related to e.g. specific mode of action of a compound allowing regulators to conclude that any potential ED related effects would be secondary to other toxicities (i.e. acetylcholinesterase activity measurements in fish in regards of neurotoxicity). Furthermore, to avoid unnecessary repetition of testing, test guidelines of *in vivo* studies, should contain all possible biological endpoints mentioned in current TGs. This would mean e.g. the inclusion of gonad histopathology in FSDT (OECD TG 234) and FSTRA (OECD TG 229).

Better definition of test concentration range. In particular, a better definition of the Maximum Tolerated Concentration (as defined in the ECHA Guidance on the application of CLP criteria (ECHA, 2024)). This would avoid testing at unnecessarily high concentrations leading to excessive toxicity (e.g. mortality or severe suffering). This would also help to better discern ED mediated effects and effects related to other toxicities. In a similar way, low concentrations which are not sufficient to conclude on the lack of potential ED-related activities or adversity should be not tested to avoid the need of multiple repetitions of the test. The development of a document giving guidance on how to best define the concentration range to test ED properties is recommended.

Foster the validation of modified protocols overcoming some of the main issues identified in available standard tests, e.g., Extended Amphibian Metamorphosis Assay (EAMA) (Ortego et al., 2021), Larval Amphibian Toxicity Test (LATT), inclusion of thyroid endpoints in fish embryo acute toxicity test (OECD TG 236). Both EAMA and LATT assays generate data on apical endpoints which are sufficient to evaluate T-mediated adverse effects that are potentially relevant at the population level. Their validation may result in the obsolescence of Larval Amphibian Growth and Development Assay (LAGDA, OECD TG 241).

Currently under validation, the proposal of the OECD project "Inclusion of thyroid endpoints in OECD fish TGs" consists of the addition of assessment of posterior swim bladder inflation, thyroid hormone levels, eye development, thyroid follicle morphology. It is still unclear how the inclusion of some of these endpoints may allow the collection of information useful to predict adversity due to endocrine (T-modality) and non-endocrine mechanisms in individuals starting from the embryonic life stage. Since those testing methods employ animals, this action point is considered to represent a short-term recommendation, although this does not reflect the potential foreseen date of validation according to OECD GD 34.

Improvement and better use of the existing OECD conceptual framework (CF) proposed in ECHA/EFSA guidance (2018) considering it as toolbox instead of a tiered approach only. This would include consideration of whether it would be possible to use non-animal approaches to interrogate observed findings and further explore mechanisms following OECD CF level 3 *in vivo* studies. Uncertainty around

apparent activity indicated by an OECD CF level 3 test could be verified with an appropriate test at lower level, before resorting to higher level tests with concomitant animal use to address adversity. This could be facilitated by developing Integrated Approaches to Testing and Assessment (IATA) which provide a framework of available methods and recommendations on best methods depending on the specific case under assessment.

Short-term recommendations

Investigate whether the current human/mammalian in vitro standard methods are sufficiently predictive of EATS-related endocrine activity in non-mammalian species and identify sufficient in vitro datasets to support a conclusion on the absence of EAS-related endocrine activity.
Currently, studies placed at the OECD CF level 2 for the investigation of endocrine activity, mainly through EAS-modalities, are specific for providing mechanistic information related to the assessment of endocrine disruption for human health, since they use either human or mammalian material. To date, no information is available on whether those can be considered predictive of possible endocrine activity in non-mammalian species. Therefore, a 2-step approach is proposed:

A. **Investigate extrapolation of findings from human/mammalian cell-based assays to non-mammalian species** through e.g. a meta-analysis comparing outcomes from *in vitro* test battery and OECD CF level 3 studies with fish, e.g. test done according to OECD TG 250, 251, 252, 229, 230.
B. **Review the sufficiency of the *in vitro* dataset for EAS-related endocrine activities** as defined by ECHA/EFSA Guidance (2018) **with the aim of supporting the conclusion on the absence of endocrine activity** based on data derived from in vitro and non-animal methodologies, as already done in human health assessment by the use of the ToxCast ER model to conclude on the E modality without the need to perform a study in line with OECD TG 440.

Better understanding of how alternatives to animal testing already in use and taken up in the regulatory world are performing in the real world.
A mapping exercise is advised to better understand identifiable mechanisms and applicability domains of recently adopted OECD TGs (e.g. RADAR and REACTIV assays). The performance of the XETA (OECD TG 248) with non-PPP compounds should also be evaluated.

Development of guidance on the use of AOPs framework to map alternatives to animal testing.
Recently, there has been a growing focus on Adverse Outcome Pathways (AOPs) as a framework for organizing mechanistic data, including data generated by alternatives to animal testing, to predict or explain adverse outcomes and the biological plausible link between endocrine activity and adversity and more generally to facilitate decision-making regarding the safety of chemicals. However, a single AOP is often insufficient to fully explain an Adverse Outcome (AO), and a network approach may be required to capture the complexity of the underlying biology. Therefore, while AOPs are valuable for the development of alternatives to animal testing, guidance on how to integrate individual non-animal methods within the AOP framework is needed to identify gaps in the methodology and inform its further development.

Mid-term recommendations

Build solid evidence and gain further knowledge on cross-species extrapolation to reduce uncertainty linked to the lack of data of relevant species.

Extrapolating biological data across species is a crucial aspect of biomedical research, pharmaceutical development and toxicology. This knowledge advancement in the field of human health has highlighted the potential to use similar methodologies to assess chemical risk across animal models used in ecotoxicology and for the prediction in wildlife. Building a solid case for the application of *in silico* cross-species extrapolation, starting from a data rich chemical/species, would create a common ground for both the scientific and the regulatory relevance. It is however crucial to gain further knowledge on structural conservation of molecular key events (KEs) and KE relationships of an AOP of relevant species. This would allow to extend the biologically plausible taxonomic domain of an identified AOP. Simultaneously, the use of cross-species extrapolation based on conservation of (part of) pathways amongst species should be seen as part of a weight of evidence approach for safety decision making within the regulatory context, and as support for the development of new relevant non-animal approaches with high potential for regulatory acceptance.

Identify pathways other than EATS and develop alternatives to animal testing and AOP for those non-EATS pathways, so that it is possible to distinguish endocrine from non-endocrine pathways.

Identify pathways other than EATS is beneficial to different aims, 1) improve the overall toxicological assessment of EDs, taking into consideration new endocrine pathways and excluding non-endocrine pathways (e.g. distinguish systemic non-endocrine specific from endocrine-specific toxicity).

Long-term recommendations

Define mechanistic based panel of in silico and in vitro assays that allows for a hypothesis driven safety assessment

A hypothesis-based safety assessment focuses on the potential bioactivity of chemicals and on key adverse outcome pathways (AOPs) relevant for environmental species. There are several benefits of this concept as recognized by pharmaceutical companies for human safety applications (Bowes et al., 2012). However, to be suitable for environmental safety, substantial work would be needed to identify critical biological characteristics in keystone species that, if impaired, would affect population levels (a primary goal within environmental safety assessment). Additionally, robust *in silico* and *in vitro* assays able to quantify those critical biological characteristics need to be developed. Those assays could be employed in the first tier of safety assessment, emulating human safety processes and potentially leveraging synergies with existing approaches in the human toxicology field.

Develop quantitative Adverse Outcome Pathways (qAOPs)

The AOP concept was developed to organize available information describing how the interaction of a chemical with a molecular target (Molecular Initiating Event, MIE) can result in an adverse outcome (AO) in individuals and, by extension, in a population. In an AOP, the MIE is causally related to the AO through a series of KEs linked by key-event relationships (KERs). As they are based on a biologically plausible suite of KEs, AOPs can be used in weight-of-evidence approaches to identify endocrine disruptors. An AOP that is sufficiently populated with substance-specific data allows establishing plausible links between KEs leading to an individual-level AO. However, both potency and exposure are required for a KE to activate the next step in the chain, that is, to reach the "tipping point." Many factors can interfere with KERs, so that an AO may not occur even if a MIE has been identified and effects on KEs are empirically observed. The integration of qAOP and TK-TD information could solve this issue (Perkins et al., 2019). Additionally,

development of qAOP networks would help in understanding extrapolation across species. Detailed qAOPs with established potency thresholds for each KE and AO prediction can however be technically challenging and resource intensive.

General recommendations

Building trust on in vitro and in silico testing.

Following scientific progress, non-animal approaches are becoming increasingly used by industry and consequently regulators are receiving more data from studies using non-animal approaches, which may not be validated methods because of lack of resources. The evaluation of these studies by a regulator can be a time-consuming process, based on the level of details and transparency of reporting. Moreover, the complexity of different approaches may have an impact on resources. Addressing and increasing understanding in both scientific and regulatory defined questions such as ADME processes, applicability domain, actual exposure quantification and therefore decreasing the overall uncertainties would qualify the method for regulatory use.

Beside this, maximising the effort to demonstrate the performance of new methods compared to *in vivo* methods (i.e. eleutheroembryos assays in parallel to Fish Short Term Reproduction Assay (OECD TG 229)) would strengthen confidence.

Overview of the available and established methodologies using non-protected life stage animal for assessing endocrine disrupting potential of chemicals

Considering the need to demonstrate activity, adversity and a plausible linkage, it appears unrealistic that a single assay can provide sufficient evidence to identify a chemical as ED, according to the criteria. Furthermore, in some cases even when *in vivo* data are available, the following considerations may limit the straightforward interpretation of standard (eco)toxicological tests: the animal model, the life stages tested, the relevance and specificity of the measured endpoints for endocrine activity/adversity, or the exposure design. Currently, *in vitro* methods are often used as supplementary or supportive information to the results of animal testing, highlighting the issue of comprehensiveness of animal testing, and the need for strategies based on integration of data from multiple test methods.

***In vivo* assays utilizing organisms at non-protected life stages according to Directive 2010/63/EU**

Available alternatives for addressing the requirement to sufficiently investigate endocrine activity in non-mammalian species are assays using eleutheroembryos at stages considered in the legislation not to be protected organisms. These mechanistic OECD CF level 3 assays do not investigate adverse effects (e.g. impaired growth and development, reduced number of offsprings). Currently there are five methods using either amphibian or fish eleutheroembryos *: Xenopus* Eleutheroembryonic Thyroid Assay (XETA, OECD TG 248), measurement of intrafollicular thyroxine (T4) using zebrafish eleutheroembryos (ZETA, non-standardized), Rapid Androgen Disruption Activity Reporter assay (RADAR assay, OECD TG 251), Detection of Endocrine Active Substances, acting through estrogen receptors, using transgenic tg(cyp19a1b:GFP) Zebrafish embrYos (EASZY, OECD TG 250), Rapid Estrogen ACTivity *In vivo* assay (REACTIV assay, OECD TG 252). The first two assays are relevant for the thyroid modality while the last three assays can be used for EAS modalities (Annex 1, Table 2). These methods are considered alternatives to vertebrate testing because eleutheroembryos used in the different assays are not yet feeding independently (their development and growth still depends on the maternal egg yolk present in their

digestive tract) and thus, they are not considered protected life stages according to EU Directive 2010/63 in the EU. However, it is not always evident that such assays end before the eleutheroembryos start free feeding. It has been observed that eleutheroembryos may start independent feeding earlier due to different testing conditions such as temperature.

In vitro assays

OECD CF Level 2 *in vitro* assays provide information on the mechanisms through which a substance could be considered endocrine active (e.g. by binding to and activating a receptor or interfering with hormone production). Some are listed in the ECHA/EFSA guidance (2018) (Annex 1, Figure 2) and others are under development and validation following the steps outlined in OECD GD 34 (Annex 1, Table 2). The *in vitro* assays are largely based on human cell test systems and as yet there is no validated test guideline for *in vitro* assays based on fish or amphibian cells for endocrine activity.

Other methodologies or ongoing initiatives with potential regulatory relevance

Other in vitro assays

Some fish and amphibian *in vitro* assays have been developed for use in the Japanese regulatory programmes EXTEND (e.g., Medaka estrogen receptor α reporter gene assay and Medaka androgen receptor β reporter gene assay for estrogenic, anti-estrogenic and androgenic activities; Medaka androgen receptor β reporter gene assay for anti-androgenic activity; *Xenopus tropicalis* thyroid receptor β reporter gene assay for thyroid and anti-thyroid activity).

Inclusion of endpoints in existing fish TGs and development of new assay with endpoints sensitive to thyroid disruptors

The ongoing OECD project 2.64 was launched in 2023 with the objective of validating thyroid morphology/histopathology, swim bladder inflation, eye development/histopathology, and thyroid hormone levels as additional endpoints in OECD TGs 236 and 210. However, swim bladder inflation and eye development/histopathology are endpoints sensitive to, but not diagnostic of, the thyroid modality, meaning that they are not specific for thyroid disruptors (Jaka et al., 2023). The inclusion of (some of) these endpoints in the zebrafish eleutheroembryo test (FET; OECD TG 236) would be relevant for thyroid-mediated activity whereas it is still not clear how measurements of such endpoints in the fish early life-stage test (OECD TG 210) could potentially provide information of adverse effects with population relevance.

As the fish early life-stage test is a common data requirement for aquatic risk assessment, the inclusion of thyroid-related endpoints in the fish early life-stage test is expected to provide information on the thyroid modality, thereby potentially reducing the need to conduct additional animal studies on amphibians. It might, however, increase the need of *in vivo* testing (for fish) under those areas of legislation that currently do not require a fish early life stage test or see it as optional. In addition, it needs to be underlined that using an *in vivo* test, including those based on eleutheroembryos, are considered as an intermediate solution and only if it reduces the overall need for animal testing until a full replacement is possible.

Similar to the transgenic fish eleutheroembryonic assays developed to detect EAS-mediated activity (Annex 1, Table 2), transgenic eleutheroembryonic assays were recently developed to investigated thyroid-mediated activity in zebrafish (Jaka et al., 2023) and Japanese medaka (Pesce et al., 2024; 2023). These assays are commercially available but have not yet undergone validation according to OECD GD34 and introduced as OECD TG.

Current efforts for integrating these alternatives in the EU regulatory frameworks

For non-target organisms other than mammals, to consider the T-mediated endocrine activity sufficiently investigated, an Amphibian Metamorphosis Assay (AMA; (OECD TG 231) should be available according to ECHA/EFSA guidance (2018). Following the publication of the OECD TG 248 (XETA) in 2019, an annex (Annex A) was added to the ECHA/EFSA Guidance (2018) explaining in which circumstances the XETA can replace the AMA for having the thyroid-mediated activity sufficiently investigated in non-target vertebrates other than mammals. This shows that the XETA can be implemented in testing strategies to inform on certain thyroid mechanisms. Mechanisms not covered by the XETA, in particular early phases of thyroid hormones synthesis, could be covered by data from *in vitro* assays (Lagadic et al., 2024) (Annex 1, Table 2). Based upon the estimates for pesticide active substances, the XETA was used for 61% of 56 substances evaluated for the thyroid modality. An AMA was conducted as a follow-up study for approximatively 10% of the substances tested in the XETA. Overall, this represents 31 AMAs avoided. Considering that a typical AMA requires approximatively 400 animals, the implementation of the XETA as defined in the ECHA/EFSA Annex A saved 12,400 animals for the evaluated pesticide active substances (Du Pasquier et al., 2025). Up to now, there is no proposal from EFSA/ECHA for using fish eleutheroembryonic assays as an alternative to the recommended OECD CF level 3 test with intact fish, e.g. the FSTRA, for EAS modalities. The applicability domains of the available fish eleutheroembryonic assays need to be further investigated especially with respect to the steps they are able to cover within the steroidogenesis pathway. Some information on this aspect will be available in the EFSA technical report on recurring issue in the context of ED assessments (in preparation). Nevertheless, regulators are recommending the use of the RADAR assay in case of specific concerns related to potential anti-androgenic properties of chemicals (Panter et al., 2023).

Main obstacles and critical aspects

Limited confidence in new methods

In the case of eleutheroembryos assays, the current limited confidence is mainly related to the use of not fully developed animals which may lead to (different) uncertainties compared to studies with more developed or adult organisms, in e.g. the AMA and the FSTRA. There is indeed evidence of not fully functional endocrine systems (e.g., thyroid hormone synthesis in the eleutheroembryonic stages used in the XETA) or of unknowns or limited metabolic competencies. This is also the case for *in vitro* test systems where metabolic competencies may be unclear These issues can be addressed by combining different methods. For example, the XETA can be combined with *in vitro* thyroperoxidase and sodium-iodide symporter inhibition assays. Similarly, the RADAR and REACTIV assays can be associated with the *in vitro* aromatase inhibition assay (OPPTS. 890.1200) and H295R steroidogenesis assay (OECD TG 456). More confidence can also be gained from comparative studies between eleutheroembryonic assays and chronic amphibian and fish assays (Du Pasquier et al., 2025; 2024). If we consider cell based/molecular assays, ADME aspects are generally lacking. There are different methods that aim to incorporate ADME in toxicity assays. If we consider fish embryos, the chorion could represent an effective barrier to chemical exposure, potentially leading to an unrealistic scenario.

Importance of the Weight of Evidence

Since the ED assessment is a complex assessment, it is important to stress that one-to-one replacement is unrealistic. To this respect, the importance of the weight of evidence, where all evidence is used and considered together with its uncertainties, is of pivotal importance. Work will be needed to better understand and provide guidance on Weight of Evidence application particularly in situations where there may be batteries of assays incorporating many alternatives to animal testing (e.g. for thyroid).

Understanding of the applicability domains of available assays

In the case of the fish eleutheroembryonic assays, these have been validated for the detection of E and A activity. While the RADAR and REACTIV assays have been shown to be able to detect effects of downstream steroidogenesis (inhibition of 5a-reductase and aromatase), it is unclear whether they are able to consistently do so for a wider range of chemicals. Simultaneously, the lack of validation and confirmation of coverage for upstream steroidogenic effects does currently not allow for these assays to cover steroidogenic activity in fish. Consequently, substitution of fish assays such as the 21-day Fish Assay (OECD TG 230) or the FSTRA (OECD TG 229) with eleutheroembryonic assays is currently not possible.

Difficult to test substances/multi constituents

The applicability domain of the currently available *in vitro* assays is unlikely to cover the variety of chemical structures and properties found in industrial chemicals. Experience based on other *in vitro* study types required under the current REACH information requirements (e.g., for skin sensitisation) has found that methods are not applicable to poorly soluble substances, metals, salts, and substances of Unknown or Variable compositions, Complex reaction products and biological materials (UVCBs). Substances with specific physico-chemical properties, such as volatility or poor solubility, can lead to false positives or render *in vitro* studies unfeasible. Volatile chemicals may result in significant chemical losses to air or contamination in neighbouring wells when using standard test systems like multi-well plates. Adaptations to these systems are necessary to ensure robust and reliable test results. However, many of those issues are not specific to *in vitro* methodologies and are applicable to *in vivo* assays (Burden et al., 2024). Additionally, a significant proportion of registered substances under REACH (up to 30%) are multi-constituent substances and UVCBs (ECHA, 2017), which pose an extreme challenge for (eco)toxicity testing.

Population relevance of endocrine-mediated adverse effects

According to Commission Regulation 605/2018, a substance is considered to be and ED if the 3 criteria are met, unless there is evidence demonstrating that the adverse effects are not relevant at the level of population. Generally, effects on growth, development and reproduction in single species in laboratory studies are regarded relevant for the maintenance of the wild population. Therefore, when effects are observed in those parameters the relevance at the level of population is inferred unless the contrary is proven. According to the ECHA/EFSA guidance, field studies/monitoring and population modelling are possible approaches for further demonstrating whether an endocrine-mediated effect is relevant at the level of population level assessment of endocrine-mediated adverse effects. However, considering the limitations of *in natura* approaches, population models appear as the most suitable tools to address population relevance (Hazlerigg et al., 2025; Crane et al., 2019). The available population models use apical endpoints (e.g. fecundity, sex ratio, time-to-metamorphosis, sexual behaviour) measured in laboratory animal studies as entries to predict population dynamics in terms of abundance and biomass. If they cannot be obtained from animal studies, these endpoints need to be generated using *in vitro-in vivo* extrapolation methods that are accurate enough to fulfil the requirements and criteria of good modelling practices. Recently a workshop was organised under the umbrella of the Society of Environmental Toxicology and Chemistry, the MAPPED[55], on the use of models to address the population relevance of endocrine-mediated effects in fish and amphibians. Publications should become available soon. [12]

[55] https://www.setac.org/resource/update-from-the-inaugural-mapped-workshop.html

References

Bowes, J., et al., Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. Nature reviews Drug discovery, 2012. 11(12): p. 909-922.

Burden, N., et al., Resource and animal use implications of the proposed REACH information requirements for endocrine disruptor assessment. Regulatory Toxicology and Pharmacology, 2024. 151: p. 105671.

Crane, M., et al., Assessing the population relevance of endocrine-disrupting effects for nontarget vertebrates exposed to plant protection products. Integrated Environmental Assessment and Management, 2019. 15(2): p. 278-291.

Development, O.f.E.C.-o.a., Test No. 456: H295R Steroidogenesis Assay, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris, 2025.

Du Pasquier, D., et al., How the Xenopus eleutheroembryonic thyroid assay compares to the amphibian metamorphosis assay for detecting thyroid active chemicals. Regulatory Toxicology and Pharmacology, 2024. 149: p. 105619.

Du Pasquier, D., et al., The Xenopus Eleutheroembryo Thyroid Assay (XETA), a New Approach Methodology (NAM) for the regulatory assessment of thyroid activity of chemicals. Environmental Toxicology and Chemistry, 2025: p. vgaf074.

European Chemical Agency and European Food Safety Authority with the technical support of the Joint Research Centre, Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009. EFSA Journal, 2018. 16(6): p. e05311.

European Chemicals Agency, Guidance on the application of the CLP Criteria – Guidance to Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures. Part 3, Health hazards. 2024: European Chemicals Agency.

European Chemicals Agency, Read-Across Assessment Framework (RAAF). 2017: European Chemicals Agency.

European Commission, Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council. Official Journal of the European Union, 2017.

European Commission, Commission Delegated Regulation (EU) 2023/707 of 19 December 2022 amending Regulation (EC) No 1272/2008 as regards hazard classes and criteria for the classification, labelling and packaging of substances and mixtures Official Journal of the European Union, 2023.

European Commission, Commission Implementing Decision (EU) 2020/569 of 16 April 2020 establishing a common format and information content for the submission of the information to be reported by Member States pursuant to Directive 2010/63/EU of the European Parliament and of the Council on the protection of animals used for scientific purposes and repealing Commission Implementing Decision 2012/707/EU (notified under document C(2020) 2179). Publication Office of the European Union, 2020.

European Commission, Commission Regulation (EU) 2018/605 of 19 April 2018 amending Annex II to Regulation (EC) No 1107/2009 by setting out scientific criteria for the determination of endocrine disrupting properties Official Journal of the European Union, 2018.

European Commission, Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. Official Journal of the European Union, 2013.

European Medicines Agency, Committee for Medicinal Products for Human Use, Guideline on the environmental risk assessment of medicinal products for human use. . 2024.

European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP), Overview of comments received on the 'Guideline on the environmental risk assessment of medicinal products for human use' (EMEA/CHMP/SWP/4447/00 Rev. 1). 2024.

European Parliament and Council of the European Union, Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. Official Journal of the European Union

European Parliament and Council of the European Union, Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Official Journal of the European Union, 2006.

Hazlerigg, C.R., et al., Integrating population-level effects into the regulatory assessment of endocrine disrupting substances. Integrated Environmental Assessment and Management, 2025. 21(3): p. 639-648.
Jaka, O., et al., Screening for chemicals with thyroid hormone-disrupting effects using zebrafish embryo. Reproductive Toxicology, 2023. 121: p. 108463.
Lagadic, L., et al., Endocrine disruption assessment in aquatic vertebrates–Identification of substance-induced thyroid-mediated effect patterns. Environment International, 2024. 191: p. 108918.
Organisation for Economic Co-operation and Development, Test No. 210: Fish, Early-life Stage Toxicity Test, OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris, 2025.
Organisation for Economic Co-operation and Development, Test No. 229: Fish Short Term Reproduction Assay, OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris, 2012.
Organisation for Economic Co-operation and Development, Test No. 230: 21-day Fish Assay: A Short-Term Screening for Oestrogenic and Androgenic Activity, and Aromatase Inhibition, OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris, 2009.
Organisation for Economic Co-operation and Development, Test No. 231: Amphibian Metamorphosis Assay, OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris, 2009.
Organisation for Economic Co-operation and Development, Test No. 234: Fish Sexual Development Test, OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris, 2011.
Organisation for Economic Co-operation and Development, Test No. 236: Fish Embryo Acute Toxicity (FET) Test, OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris, 2025.
Organisation for Economic Co-operation and Development, Test No. 240: Medaka Extended One Generation Reproduction Test (MEOGRT), OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris, 2023.
Organisation for Economic Co-operation and Development, Test No. 241: The Larval Amphibian Growth and Development Assay (LAGDA), OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris, 2015.
Organisation for Economic Co-operation and Development, Test No. 248: Xenopus Eleutheroembryonic Thyroid Assay (XETA), OECD Guidelines for the Testing of Chemicals, Section 2. 2019.
Organisation for Economic Co-operation and Development, Test No. 250: EASZY assay - Detection of Endocrine Active Substances, acting through estrogen receptors, using transgenic tg(cyp19a1b:GFP) Zebrafish embrYos, OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris, 2021.
Organisation for Economic Co-operation and Development, Test No. 251: Rapid Androgen Disruption Activity Reporter (RADAR) assay, OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris, 2022.
Organisation for Economic Co-operation and Development, Test No. 252: Rapid Estrogen Activity In vivo (REACTIV) assay, OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris, 2024.
Organisation for Economic Co-operation and Development, Test No. 440: Uterotrophic Bioassay in Rodents: A short-term screening test for oestrogenic properties, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris, 2007.
Ortego, L.S., et al., The extended amphibian metamorphosis assay: A thyroid-specific and less animal-intensive alternative to the larval amphibian growth and development assay. Environmental Toxicology and Chemistry, 2021. 40(8): p. 2135-2144.
Panter, G.H., et al., Detection of anti-androgenic activity of chemicals in fish studies: a data review. Critical Reviews in Toxicology, 2023. 53(5): p. 326-338.
Perkins, E.J., et al., Building and applying quantitative adverse outcome pathway models for chemical hazard and risk assessment. Environmental toxicology and chemistry, 2019. 38(9): p. 1850-1865.
Pesce, E., et al., A novel transgenic model to study thyroid axis activity in early life stage medaka. Environmental Science & Technology, 2023. 58(1): p. 99-109.
Pesce, E., et al., Testing the sensitivity of the medaka Transgenic Eleuthero-embryonic THYroid-Specific assay (TETHYS) to different mechanisms of action. Aquatic Toxicology, 2024. 276: p. 107081.
Solecki, R., et al., Scientific principles for the identification of endocrine-disrupting chemicals: a consensus statement. Archives of Toxicology, 2017. 91(2): p. 1001-1006.
United States Environmental Protection Agency, Endocrine Disruptor Screening Program Test Guidelines OPPTS 890.1200: Aromatase (Human Recombinant). 2009.
World Health Organization, Global assessment of the state-of-the-science of endocrine disruptors. International Program on Chemical Safety, 2002.

3.5. Birds and mammals

Introduction

The document outlines current environmental safety assessment (ESA) approaches using birds and mammals which are relevant to EU chemical safety regulations. It draws mostly from pesticide legislation, which has the highest demand for terrestrial vertebrate testing. To register a pesticide, applicants typically generate data using around 5,500 laboratory birds and re-use data from approximately 5,000 laboratory mammals in the human safety studies for the terrestrial vertebrate risk assessments (see details in Table 5).

Field studies using wild animals can be requested on a case-by-case basis, typically employing low severity and non-lethal assessments, similar to conservation research (e.g., radio-tracking). Such studies are unlikely to be requested under some regulatory regimes, e.g. the REACH regulation. However, these studies are not mandatory and fall outside the scope of this document and thus roadmap. Existing data can also be used within Weight of Evidence.

This document highlights current data requirements for laboratory in-vivo studies, as well as proposals for short-, mid- and longer-term action, to arrive at a new destination, where regulatory-relevant decisions can be made using data generated from best available non-animal-based scientific tools.

Terrestrial wildlife is part of safety assessments in several chemicals risk assessment regulations, including plants, microorganisms and invertebrates. However, these do not trigger the use of vertebrates unless a chemical product is intended for use in habitats they occupy—for example, birds and mammals in areas where pesticide or biocide products may be applied.

The most stringent terrestrial wildlife chemical safety assessment in Europe is in the pesticide legislation, Regulation (EU) No 1107/2009. The data requirements for terrestrial safety testing for pesticidal active substances are described in Regulation (EU) No 283/2013.

In vivo testing requirements have been considered across all five vertebrate animal classes. For mammal safety, a wide range of studies are performed to investigate human safety across different life stages (in utero to breeding), exposure durations (acute, chronic, lifespan) and exposure routes (oral, dermal or inhalation). These studies also consider interspecies concordance, typically involving rodents, rabbits, dogs, and occasionally other species. The data generated from these laboratory mammal studies are re-used as the foundation for assessing environmental safety of terrestrial vertebrates, including wild mammals. No additional laboratory mammal testing is required specifically for ecological effect assessments of pesticides.

The only additional tests with terrestrial vertebrates required for ecological effect testing for pesticides are those with birds. Both acute and chronic toxicity data are required for at least one avian species as part of pesticide authorization/registration in Europe. For global registration of a pesticide, further acute or chronic testing may be triggered depending on regional requirements. In most cases, the number of laboratory birds used for ESA can be comparable to the number of laboratory mammals used for human safety assessments for a pesticide, though this equivalence is not widely acknowledged (Table 5).

No additional laboratory birds or mammals are used for bioaccumulation testing, nor are they required for ecological endocrine screening.

Extrapolation of effect information between animal classes or within species is currently feasible on a case-by-case basis. However, our understanding of the physiological, anatomical, and biochemical differences among terrestrial vertebrates—particularly in terms of absorption, metabolism, toxicokinetics, and toxicodynamics—remains limited. Advancing research in these areas could support the development of more regionally tailored risk assessments, potentially accounting for local breeding patterns and migration behaviours.

Bird and mammal *in vivo* data are used to determine relevant endpoint values used in acute or chronic risk assessments. Under Regulation (EU) No 1107/2009, the risk assessment outcome for terrestrial vertebrates is based on a toxicity-exposure ratio (TER), which uses results from *in vivo* studies to establish the initial toxicity value—typically, the No Adverse Effect Level (NOAEL). One possibility to refine the risk could be to consider the ecological (population) relevance of laboratory observations or by applying mechanistic effects knowledge, such as those derived from toxicokinetic-toxicodynamic (TKTD) models or other scientifically justified approaches, although currently they are not widely used and/or accepted. A TER above 10 for acute assessments is considered protective against visible mortality, while a TER above 5 for chronic assessments is assumed to safeguard against long-term impacts on population abundance and biodiversity. In chronic scenarios, the population is regarded as the ecological entity of concern.

Beyond pesticide ERAs, data from laboratory or wild birds and mammals may occasionally be available for ESAs under REACH. However, such data is not available in the majority of cases. The REACH regulation requires ERA for predators, including birds and mammals, exposed via the terrestrial and aquatic food chains, referred to as 'secondary poisoning', in cases where a substance has potential to bioaccumulate and also a potential to cause toxic effects if accumulated in higher organisms. Usually, only mammalian toxicity information available from the human health risk assessment (not discussed further here) is used for such assessments. It is acknowledged in REACH guidance (Chapter R7c) that birds are fundamentally different from mammals in certain aspects of their physiology (e.g. the control of sexual differentiation, egg laying, etc.), and so mammalian toxicity data are of limited predictive value for birds.

Bird and mammal testing specifically for ERA is not expected under EU regulations governing cosmetics, pharmaceuticals, or veterinary medicines. The use of animals for efficacy testing of regulated chemicals—such as human or veterinary medicines—is considered outside the scope of this document. In addition, testing on target species is outside of the scope of the roadmap.

A successful alternative test method should not require any new animal material—such as primary cells, plasma, embryos, or unprotected life stages. It should, either alone or in combination with other tests, provide a level of protection that is equal to or greater than that provided by current approaches, and be suitable for regulatory decision-making.

To be considered successful, the method must undergo suitable validation steps. This includes demonstrating that its biological and chemical applicability domains are well-defined, and that it performs reliably and reproducibly under testing conditions in a laboratory setting.

While population-level effects are the ecological entity of concern, most *in vivo* studies primarily capture effects at the level of individual animals. To bridge this gap, computational models are already being used to support population-level effect assessments on a case-by-case basis. However, these models still face challenges in validation and currently have limited regulatory acceptance. Ongoing projects initiatives—such as the PARC (Partnership for Assessment of Risk of Chemicals), SETAC Effect modelling interest group[56] and the EFSA WG on effects models in ERA[57]—are actively working to advance the development, validation, and regulatory uptake of these tools.

Progress in developing animal-free methods to predict individual-level effects varies across the five vertebrate classes. Significant advancements have been made in mammalian safety testing, largely driven by its relevance to human safety. In contrast, avian safety testing methods have seen limited innovation since the 1980s—though this is beginning to change. Within PARC various models are being developed for in total 18 domestic and wildlife birds (publication, in preparation), which might be used in screening and und in a weight of evidence, this document focuses specifically on these two terrestrial vertebrate classes.

[56] SETAC Working group: Model Acceptability and scenario Development (MAD) (https://www.setac.org/group/effect-modeling/document.html?id=0332F6DE-F1F3-4621-A395E98FE6A9CE84).
[57] EFSA WG on effects models in ERA. Accessed via: https://open.efsa.europa.eu/working-group/300000037091043

Wild mammals

For wild mammals (non-humans), individual-level effect assessments typically begin with the mammalian dataset originally generated for human safety evaluations. These datasets currently include *in vivo* studies addressing acute, chronic, developmental, and reproductive outcomes. As *in vivo* tests are progressively replaced by non-animal alternatives, it is essential to carefully assess the ecological relevance of these replacement(s).

Currently, no mammalian assays are in development specifically for ecological purposes that are not also intended for human safety assessment. However, assays developed for human health but later deemed to have limited human relevance may still hold value for ecological assessments. For example, the mouse embryonic stem cell test (Seiler & Spielmann, 2011) may be considered less human-relevant than its human-derived counterpart (Gabdoulline et al., 2015), yet it could still be informative for ESA.

To support progress in this area, continued collaboration between stakeholders focused on non-animal approaches for human safety and those working on wild vertebrate protection is strongly recommended.

Wild birds

Laboratory bird studies on wild birds typically investigate acute, chronic or reproductive outcomes. These studies are most commonly performed using the mallard duck (*Anas platyrhynchos*) and a quail (*Coturnix japonica* or *Colinus virginianus*), with other species such as canaries occasionally required outside Europe. As *in vivo* tests are gradually replaced by non-animal alternatives, it is essential to evaluate the ecological relevance of these replacement(s).

The core *in vivo* test methods for avian toxicity have remained largely unchanged since their initial publication in the 1980s. While protocols for acute testing were updated in OECD Test Guideline 223 (2016), they still rely on purpose-bred birds that do not survive the study. Currently, no non-animal studies are accepted by global regulatory authorities—excluding non-testing approaches such as read-across or data waiving, if agreed. No in vitro or in silico assays are yet sufficiently advanced to serve as short-term replacements for existing *in vivo* models.

However, promising developments are emerging:

- Commercial availability of avian cell line, such as chicken and quail for non-biomedical uses[58]
- Development of qPCR arrays using quail embryo hepatocytes (animal-based) (Crump et al., 2023)
- Publication of the first avian physiologically based kinetics and bioenergetics models (Martin et al., 2024)
- The UK NC3Rs "Wings of Change" challenge (2024-2028), which funds the development of New Approach Methodologies (NAMs) to assess acute and chronic avian toxicity for chemical screening and environmental risk assessment[59]

It is recommended to maintain awareness of these and other avian-specific developments, and to explore their potential for regulatory use. Additionally, leveraging cross-species research and concepts—such as virtual control groups, in vitro-*in vivo* extrapolation (IVIVE), and dynamic energy budget – toxicokinetic-

[58] University of Edinburgh National Avian Research Facility https://vet.ed.ac.uk/roslin/engagement/facilities/national-avian-research-facility/avian-resources
[59] UK National Centre for the replacement, refinement & reduction of animals in research (NC3Rs) CrackIT challenge https://nc3rs.org.uk/crackit/wings-change

toxicodynamic models—can accelerate progress. To support this, effective collaboration between stakeholders focussed on non-animal approaches for all wild vertebrate safety should be actively encouraged and facilitated.

Vision

The vision centres on developing and implementing non-animal approaches capable of predicting regulatory-relevant endpoints such as the acute LD50 and the chronic NOAEL. Achieving this requires a deep understanding of the biological effects observed in traditional (*in vivo*) studies and how they inform regulatory decisions, along with strategies to replicate or improve upon them using alternative approaches.

Two complementary concepts have emerged to support this goal: the Adverse Outcome Pathway (AOP) framework, and an in vitro / *in vivo* testing strategy adapted from the human safety paradigm. These are briefly summarised below, with further details provided in the relevant Annex.

The toxicology parallelogram adapted for ecotoxicology

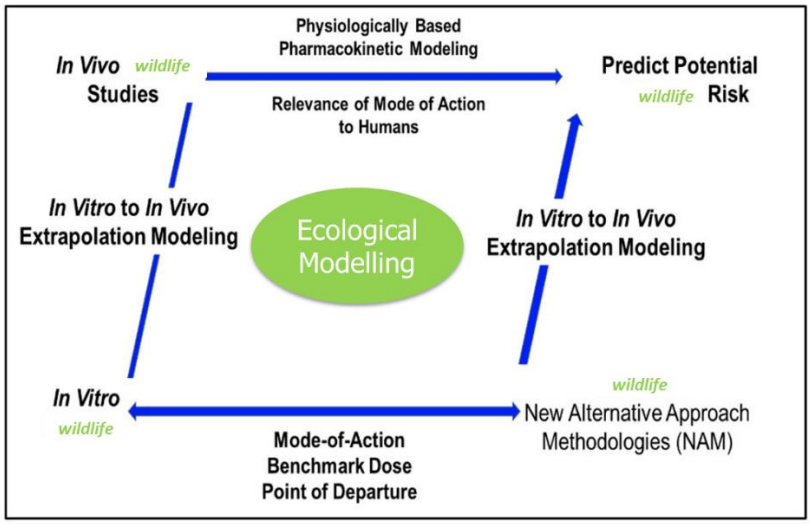


Figure 6: a parallelogram view of how *in vivo* and in vitro data combine with ecological modelling to predict potential wildlife risk.

For over a century, medical researchers and toxicologists have relied on laboratory animals as surrogates for humans, using *in vivo* animal studies to predict and extrapolate potential human health risks. Over time, the tools and methodologies available have evolved significantly.

Today, emerging in vitro assays—both animal- and human-based—combined with advanced computational modelling, offer the potential to match or even surpass the predictive power of traditional animal tests for human risk assessment. This concept is visualised as a 'parallelogram' model, illustrating the relationship between different data sources and their predictive value.

It is now proposed to adapt this parallelogram framework to illustrate how wildlife risk assessments could similarly benefit from the integration of diverse testing approaches. This adapted model would reflect the unique challenges and opportunities in predicting ecological effects using non-animal methods.

The Adverse Outcome Pathway (AOP) framework

The Adverse Outcome Pathway (AOP) framework, first introduced by Ankley et al. (2010), was designed to support the integration of emerging mechanistic data into ecological risk assessments. It provides a structured approach to link molecular-level events to adverse outcomes that are meaningful for regulatory decision-making—specifically, effects on survival, development, and reproduction in individual organisms and, ultimately, impacts at the population level.

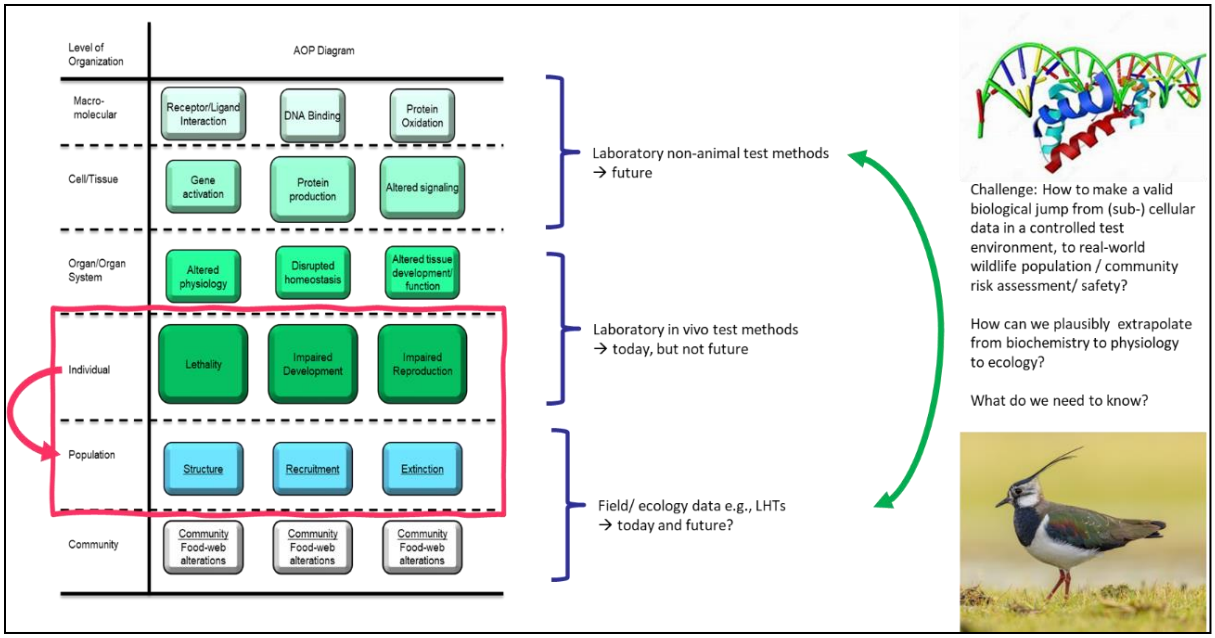


Figure 7: The Adverse Outcome Pathway (AOP) concept illustrates how data from diverse test methods can be integrated to identify adverse effects relevant for risk assessment—specifically to protect wild bird populations from known chemical stressors.

The AOP framework anticipates how information from different levels of biological or ecological complexity can be integrated to support regulatory decision-making. A central feature of this approach is the identification of individual-level endpoints that can be used to predict population-level responses through computational modelling.

For a non-animal approach to be regulatory-relevant, it must generate the type of information needed by regulators. The AOP framework helps map how various types of biological data can be linked to a sub-lethal adverse effect—such as those used to derive a NOAEL—or to lethality endpoints relevant for acute risk assessments. To optimise this approach, current scientific knowledge about toxicity mechanisms is being compiled, and knowledge gaps are being identified. This results in a plausible, qualitative, and chemically agnostic description of how different observations from *in vivo* studies are connected to adverse outcomes.

However, for a chemical-specific risk assessments, qualitative AOPs are not sufficient. Deriving a numerical value equivalent to a NOAEL requires a quantitative AOP (qAOP), which incorporates chemical specific data for key events and parameters. Although several AOPs have emerged since their introduction to ESAs over 15 years ago, no validated qAOPs are currently available.

To enable animal-free ESA, it is recommended that significant efforts be directed toward the development and validation of relevant AOPs. These should be integrated with existing and emerging ecologically relevant computational models - such as dynamic energy-TKTD modelling - to support their implementation in regulatory decision-making.

Optimising ecological realism in risk assessment

Globally, there are approximately 11,000 bird species and 6,000 mammal species, with around 600 bird species and 250 mammalian species found in Europe. Despite this diversity, current global testing requirements for chemical safety assessments typically involve only up to 4 mammalian species (rat, mouse, rabbit, dog) and one to three bird species (duck, quail, canary). Risk assessments rely on adjustment factors to extrapolate data from these few laboratory species to thousands of wild species. As ecological knowledge grows, the reliability of cross-species extrapolation continues to improve.

In addition, even within a single species, individuals may inhabit diverse landscapes and be exposed to varying stressors—chemical, climatic, or ecological. To enhance the relevance of environmental safety assessments, it is recommended to incorporate species-specific ecological information, including life history traits, population dynamics, habitat, and landscape context. This enables more accurate and regionally tailored decisions at the country, zonal, or European level.

While this recommendation may extend beyond the immediate scope of phasing out of animal testing for chemical safety assessments, it aligns with the overarching goal of the European Commission roadmap to protect wildlife using the best available science. Continued investment in ecological research and cross-species extrapolation is essential to inform these assessments and to monitor their effectiveness as non-animal test methods are adopted.

Recommendations

Currently, only a limited number of ready-to-implement animal-free approaches exist to fully replace *in vivo* testing with birds and mammals. The process of developing the roadmap has helped identify key methodological development gaps that, if addressed, could support the future phase-out of animal testing.

The main recommendation is to establish a structured, collaborative process—within the organisational framework implementing the roadmap and in partnership with stakeholders such as EPAA and the research community—to systematically work toward the goal of phasing out avian studies for regulatory purposes.

Priority actions:

- Develop AOP models: Focus on regulatory-relevant effects by building AOPs that establish causal links between molecular or cellular data and adverse outcomes. Progress toward qAOPs to enable derivation of endpoint values for risk assessment.
- Leverage advances from human safety science: Apply and adapt concepts already in development for human toxicology, including:
    - Virtual control groups,
    - In vitro-*in vivo* extrapolation (IVIVE),
    - Dynamic energy budget-toxicokinetic-toxicodynamic modelling,
    - qAOPs.

  These approaches should be applied across and between all vertebrate classes, wherever scientifically justified.

- Fund targeted research for ecological relevance: Support studies to identify key bird and mammal species for regulatory focus, including:
    - Species-specific traits and population dynamics,
    - Cross-species extrapolation methods,
    - Integration of habitat and landscape-level data to inform species prioritisation at national, zonal, or EU level

References

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, Serrrano JA, Tietge JE, Villeneuve DL. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem. 2010 Mar;29(3):730-41. doi: 10.1002/etc.34. PMID: 20821501.

Article 1 of Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32010L0063 accessed February 2025.

Crump D, Hickey G, Boulanger E, Masse A, Head JA, Hogan N, Maguire S, Xia J, Hecker M, Basu N. Development and Initial Testing of EcoToxChip, a Novel Toxicogenomics Tool for Environmental Management and Chemical Risk Assessment. Environ Toxicol Chem. 2023 Aug;42(8):1763-1771.

EFSA (European Food Safety Authority), Aagaard A, Berny P, Chaton PF, Antia AL, McVey E, Arena M, Fait G, Ippolito A, Linguadoca A, Sharp R, Theobald A and Brock T, 2023. Guidance on the risk assessment for Birds and Mammals. EFSA Journal 2023; 21(2):7790, 300 pp. https://doi.org/10.2903/j.efsa.2023.7790. https://doi.org/10.2903/j.efsa.2023.7790

EFSA. 2014. Scientific Opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products EFSA Panel on Plant Protection Products and their Residues (PPR). European Food Safety Authority (EFSA), Parma, Italy. EFSA Journal 2014;12(3):3589.

Gabdoulline R, Kaisers W, Gaspar A, Meganathan K, Doss MX, Jagtap S, Hescheler J, Sachinidis A, Schwender H. Differences in the Early Development of Human and Mouse Embryonic Stem Cells. PLoS One. 2015 Oct 16;10(10):e0140803.

Kramer VJ, Etterson MA, Hecker M, Murphy CA, Roesijadi G, Spade DJ, Spromberg JA, Wang M, Ankley GT. Adverse outcome pathways and ecological risk assessment: bridging to population-level effects. Environ Toxicol Chem. 2011 Jan;30(1):64-76. doi: 10.1002/etc.375. PMID: 20963853.

Martin T, Bauer B, Baier V, Paini A, Schaller S, Hubbard P, Ebeling M, Heckmann D, Gergs A. Reproductive toxicity in birds predicted by physiologically-based kinetics and bioenergetics modelling. Sci Total Environ. 2024 Feb 20;912:169096.

Perkins, E.J., Ashauer, R., Burgoon, L., Conolly, R., Landesmann, B., Mackay, C., Murphy, C.A., Pollesch, N., Wheeler, J.R., Zupanic, A. and Scholz, S., 2019. Building and applying quantitative adverse outcome pathway models for chemical hazard and risk assessment. Environmental toxicology and chemistry, 38(9), pp.1850-1865.

Schmolke A, Thorbek P, DeAngelis DL, Grimm V. 2010. Ecological models supporting environmental decision making: A strategy for the future. Trends Ecol Evol 25:479–486.

Scientific Opinion of the Scientific Panel on Plant Protection Products and their Residues (PPR Panel) on testing and interpretation of comparative in vitro metabolism studies | EFSA

Seiler, A., Spielmann, H. The validated embryonic stem cell test to predict embryotoxicity in vitro. Nat Protoc 6, 961–978 (2011).

The validated embryonic stem cell test to predict embryotoxicity in vitro | Nature Protocols

3.6. Ecological thresholds of toxicological concern and species sensitivity distributions

Background

Ecological thresholds of toxicological concern (ecoTTC) and species sensitivity distributions (SSDs) are both probabilistic methods that use mathematical models to identify chemical concentration thresholds below which a minimal probability of negative effects on environmental communities is expected. Deterministic hazard assessment approaches often utilized in regulatory schemes determine protective thresholds via the use of individual hazard data points (i.e., the most sensitive acute or chronic toxicity value) and uncertainty factors to generate predicted no effect concentrations (PNECs) for individual chemicals. Compared to that probabilistic methods such as ecoTTC and SSD leverage the full breadth of existing data and allow to easily integrate newly generated data.

If sufficient data for deriving them are available, ecoTTCs and SSDs (as well as other probabilistic hazard assessment approaches such as chemical toxicity distributions) are robust methods that can be used to minimize the need for animal data when assessing chemical safety. The approaches can also be used to analyse the value of generating new data (e.g., fish toxicity data) for refining the assessment (i.e., identified protective thresholds). Thus, not only can new animal testing be avoided to safely support chemical use, but the value in additional testing can be quantified. Integrating these approaches in chemical safety assessments serve as a tool to reduce animal testing.

Recommendations

Medium-term actions

- Harmonisation of methods and applied criteria, improvement of methods
    - For SSD: Harmonisation is needed on the selection and use of assessment factors in addition to the x percentile of hazardous concentrations (HCx).
    - For ecoTTC: Acceptable criteria for chemical grouping need to be defined in guidance.
    - For ecoTTC: Comparisons against reference data are needed to characterize accuracy and to determine how conservative are thresholds determined with ecoTTC.
- Better integration of approaches in the chemical safety assessment framework
    - Guidance is needed to clarify how to use the approaches in chemical safety assessment. Clarity and acceptance is needed regarding broader utilization of available data (e.g., acute toxicity data, non-standard species).
    - Probabilistic hazard assessment should be built into core tools used for chemical safety assessments such as the OECD Toolbox.

Ecological Thresholds of Toxicological Concern (EcoTTCs)

The ecoTTC is an approach for identifying a single environmental concentration below which adverse effects on environmental communities are expected to be minimal for a group of toxicologically or chemically similar substances. This method supports reduction in animal testing by leveraging existing data and minimizing the need for new ecotoxicity studies. Free software is available to support calculations of ecoTTCs.

The ecoTTC is a threshold concentration using a selected percentile (often the 5th percentile) of the probability distribution of Predicted No Effect Concentrations (PNECs) derived from acute or chronic toxicity data for similar chemicals. This is a conservative approach that identifies a hazardous concentration

threshold that is determined based on a distribution of concentrations at which no toxicity has been identified. Data leveraged are from chemicals that share relevant characteristics to the substance or chemical class of interest (e.g., mode-of-action, physicochemical properties) ensuring that the derived hazardous concentration is fit for purpose. Inclusion of data based on quality and appropriateness for grouping must be individually considered.

Because of the use of grouped data, the ecoTTC approach is valuable in providing screening-level assessments for chemicals with minimal toxicity data and may be able to provide insight for more complex assessments (e.g., UVCBs). However, it does require the availability of robust, chemical-specific hazard data with sufficient taxonomic coverage to generate high-quality PNECs. Additionally, it is likely a conservative approach given the inclusion of uncertainty factors in the derivation of PNECs. A similar, but alternative, approach that evaluates a distribution of effect concentration values without uncertainty factors – known as a chemical toxicity distribution (CTD) – might be more relevant when considering specific species or trophic levels, particularly if one trophic level is uniquely sensitive.

The ecoTTC method is not yet leveraged in regulatory frameworks for environmental assessments. Opportunities for improving robustness include the need for broader toxicity data on non-standard test organisms, better-defined criteria for chemical grouping, and validation through comparisons with reference data to characterize the method's accuracy.

Species Sensitivity Distributions (SSDs)

An SSD is an approach for representing differential sensitivity to a single toxicant across environmental species. Similar to ecoTTC, it can be leveraged to identify protective chemical environmental concentrations beneath which a minimal probability of negative effects on environmental communities would be expected. However, in this case the focus is the hazard potential of an individual chemical. If sufficient data is available, analysis can be quickly done via freely available software and open-source tools.

Protective thresholds identified using SSDs are also usually based on the 5^{th} percentile, known as the 5^{th} percentile hazardous concentration (HC5), though the probability distribution is based on effect concentration values derived from acute or chronic environmental toxicity data across various species, not PNECs. SSDs are considered to be conservative and robust approach for assessing environmental hazards of individual chemicals.

A drawback is that SSDs require a significant amount of data across several taxonomic groups to represent the potential diversity of sensitivity to a toxicant in the environment. Hence, the method will only be applicable for reduction of animal testing for those substances, for which data from several taxonomic groups are already available. Global requirements for utilizing SSDs in regulatory toxicology require vertebrate data (i.e., fish) to be included as part of the taxonomic coverage. Without consideration of data generated by non-standard species or guideline assays, an SSD can be difficult to construct. For example, an SSD cannot be generated based on data requirements for chemical registration under REACH alone.

SSDs are used in some capacity for chemical safety assessment in various global regulatory frameworks, including the EU. However, the approaches are not harmonized. Additionally, the value of effort in these higher-tier approaches is not visible in current chemical safety requirements and guidance, thus creating challenges for implementation. Clarity on broader utilization of data (e.g., acute toxicity data, non-standard species) and use of application factors in addition to HCx (often removing the value of SSD assessment versus standardized deterministic approaches) represent some of the immediate needs for increased use.

4. Recommendations and actions for human health assessments

4.1. Acute toxicity (human health assessment)

Replacement of *in vivo* testing with in-silico methods: pilot project on acute oral toxicity

Introduction

This section describes a strategy and stepwise approach for replacing *in vivo* testing in animals to determine the acute oral toxicity (AOT) of chemicals with *in silico* methods[60].

As part of the work in the Human Health Roadmap Working Group (HH WG), an initiative has been started by stakeholders including the European Chemicals Agency (ECHA), the United States National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), the German Federal Institute for Risk Assessment (BfR), the European Partnership for Alternatives to Animal Testing (EPAA) and the European Food Safety Authority (EFSA) to evaluate the feasibility of such a replacement. The strategy includes legal analysis, performance benchmarking, evaluation of existing *in silico* methods and a phased approach to regulatory integration.

As a pilot, the activity is also meant to serve as a template for similar, possibly more complex cases in the future by

- exploring the legal options for the use of *in silico* methods instead of *in vivo* testing for AOT assessment and to this end, the need for a corresponding adaptation of legal texts or guidance;
- implementing a first, selected *in silico* method for that purpose; further methods can then be implemented based on the experiences gained;
- exploring and, where necessary, proposing adaptations to the legal text and/or guidance to enable options for making *in silico* methods the default approach for AOT assessment.

Below, these steps are elaborated in more detail for two Regulations, i.e. Regulation (EC) 1272/2008 (CLP) and Regulation (EC) 1907/2006 (REACH). However, following initial discussions, the activity is meant to be extended to all other legislations in the scope of the Roadmap, which require[61] – or use the results from - AOT testing.

As the strategy will explore the use of already existing *in silico* tools, it addresses “basket” 1 (short-term actions) of the roadmap.

Approach

The stepwise approach followed in this activity can be summarised as follows:

1. Determine the generic regulatory needs and requirements with respect to AOT assessment.

[60] For this document, the term “*in silico* tool” refers to a computer model, or suite of models, available in the form of executable software and used to predict physico-chemical, toxicokinetic or toxicodynamic properties of chemicals. In contrast, the term “*in silico* method” refers to a conceptual approach for assessing such properties in the context of chemical risk assessment utilising one or more specific *in silico* tools.

[61] Of note, this activity is not applicable to the human pharmaceutical sector where conventional acute toxicity studies are no longer required (ref ICH M3(2)). For these products relevant information can be obtained from other toxicity studies carried out as part of the drug development program. Also for veterinary medicinal products it is recommended to derive toxicological reference values from other study types and waiving of acute oral toxicity studies is usually accepted. Especially data of LD_{50} studies are of limited value and are normally not accepted as single basis for quantitative risk assessment (EMA/CHMP/CVMP/3Rs/164002/2016 Rev. 1).

2. For each specific legislation: determine whether the legislation itself or associated guidance needs to be updated, or whether additional guidance needs to be developed to

   a) **allow for the replacement** of *in vivo* AOT testing with *in silico* methods in general and

   b) **make in silico methods the default option** for AOT characterisation under the specific legislation.

3. Develop **generic performance/acceptability criteria** and **benchmarks** for AOT *in silico* methods to meet in order to establish scientific and regulatory confidence in using these methods.

4. For each specific *in silico* method considered: establish **scientific confidence**, including a characterisation of the method's **applicability domain** and whether and how it meets the benchmarks under 3.

5. Formulate an **overall implementation plan** incl. action items, milestones and timelines, taking into account any open questions identified under steps 1 – 4.

Regulatory needs and requirements

In vivo AOT testing for chemical safety in the EU is currently performed acc. to OECD Test Guidelines (TGs) 420 (Acute Oral Toxicity - Fixed Dose Procedure), 423 (Acute Oral toxicity - Acute Toxic Class Method) or 425 (Acute Oral Toxicity: Up-and-Down Procedure). In addition, historical data generated by using the now obsolete OECD TG 401 (Acute Oral Toxicity) are used for AOT assessments, where available. These methods serve to determine a so-called Acute Toxicity Estimate (ATE), i.e. a (semi-)quantitative estimate of the single dose or a dose / dose range (in mg kg/bw), such as the LD50 (lethal dose for 50% of the animals tested at that dose level) characterising the acute toxicity of a substance or mixture when administered orally to rodents.

Need for updating existing legislation or guidance

CLP Regulation

Use of in silico tools for AOT assessment in general

Under CLP, in silico tools may be used for AOT classification, on condition that

- they are scientifically valid (plausible, reliable, sufficiently predictive) and
- they provide ATEs suitable for deciding on placing – or not – chemicals into one of the CLP AOT categories.

No changes of the CLP legal text appear necessary to allow for using *in silico* methods for AOT classification and labelling, provided their adequacy and reliability is demonstrated. It should be considered, however, to remove explicit (albeit non-prescriptive) reference to *in vivo* methods from the text and to add the possibility to use *in silico* tools directly.

In addition, once the adequacy and reliability of specific *in silico* methods has been agreed upon, it would seem helpful to perform (minor) updates of the CLP guidance document, explaining their applicability as well as potential caveats, if any.

Make in silico tools the default option for AOT assessment

To make the use of *in silico* methods the default approach for classification and labelling under CLP would require an update of Chapter 3.1 of the United Nations' Globally Harmonised System and the corresponding Chapter 3.1 in Annex I of the CLP Regulation.

REACH

With respect to REACH, no further requirements in addition to those already mentioned in the previous section on CLP have been identified.

Use of in silico tools for AOT assessment in general

No changes of the REACH legal text appear strictly necessary to allow for using results from *in silico methods* for AOT classification and labelling and/or risk assessment, provided their adequacy and reliability is demonstrated. It should be considered, however, to review and, where necessary, revise the respective REACH guidance document on Information Requirements/Chemical Safety Assessment (IR/CSA), Chapter R.7a, to remove explicit (albeit non-prescriptive) reference to *in vivo* methods and to add the possibility to use *in silico* tools directly.

In addition, once the adequacy and reliability of specific *in silico* methods has been agreed upon, it would seem helpful to perform a (minor) update of this part of the REACH guidance, explaining their applicability as well as potential caveats, if any.

Make in silico tools the default option for AOT assessment

To make *in silico* methods the default approach for AOT characterisation under REACH, an update of Annex VII of Regulation (EC) No 1907/2006 would be required.

Other regulations in the scope of the COM roadmap

Although the analysis of further regulations is pending, it is not expected that data from existing AOT *in vivo* tests are used in a significantly different way in other EU chemical legislations as compared to REACH and CLP.

Scientific and regulatory confidence in *in silico* methods for AOT assessment

To evaluate whether *in silico* methods can be trusted scientifically, their adequacy and reliability for providing reliable estimates of AOT has to be assessed. The QSAR Assessment Framework (QAF), published by the Organisation for Economic Co-Operation and Development (OECD) in 2023, offers a number of Assessment Elements (AEs) for this purpose, as well as for assessing the reliability of predictions made using these models.

However, the QAF itself does not provide performance benchmarks that have to be met by an *in silico* method in order to be acceptable for a given regulatory purpose.

Assuming a specific *in silico* method has been found scientifically acceptable in general, its suitability in regulatory terms[62] is mainly determined by its applicability domain (AD) and predictive performance. In the context of replacing an existing *in vivo* test method, a comparative element is added, as expressed by the following questions:

- To what extent does the AD of the *in silico* method cover (or perhaps even exceed) that of the *in vivo* method to be replaced? Are there sectors of chemical space for which the use of the *in silico* method is not possible or advisable?
- Is the predictive accuracy of the *in silico* method, i.e. the degree to which the in silico method is capable of reproducing the results of the *in vivo* test, acceptable?

The assessment of predictive performance first requires the development of generic acceptance criteria. In other words, for AOT assessment, the question has to be answered what percentage of misclassifications vs. the previous method or what degree of deviation of the predicted ATE vs. the "true" ATE, e.g. a measured

[62] In cases where a 3R's compliant methodology is considered to be used in regulatory testing in the pharmaceutical sector, the developer can apply for qualification advice or a qualification opinion through EMA's Scientific Advice Working Party. EMA also offers early interactions to NAM developers via the Innovation Task Force.

LD_{50} value, is considered acceptable. Considerations regarding e.g. what degree of deviation was accepted when introducing OECD TGs 420, 423 and 425 to replace the previous, now obsolete OECD TG 401, as well as an analysis of variability in the outcomes from multiple AOT tests available for the same substance, can help in establishing such benchmarks.

In principle, generic benchmarks developed in the context of CLP (for both, predicting ATE values/ranges or classifications) would also pertain to REACH, as far as classification and labelling are concerned. In the rare cases, where ATEs are used as PoDs for risk assessment, no additional benchmarks would seem to be required on first thought.

Confidence in specific in silico methods

As the most suitable *in silico* tool to be investigated by this pilot first, a specific *in silico* modelling suite, CATMoS, has been selected, because it

a) represents a consensus suite of a variety of models developed in different regions of the world and

b) has already undergone extensive prevalidation.

It is expected that, once generic benchmarks have been defined and the performance of CATMoS has been assessed, other available models can be added with comparatively little extra effort.

The Collaborative Acute Toxicity Modelling Suite (CATMoS)

Description of the method

Following a workshop in 2018 (Kleinstreuer et al., 2018), the U.S. National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), together with many international collaborators, has built a suite of *in silico* tools under the name of CATMoS (Collaborative Acute Toxicity Modelling Suite). CATMoS is part of the *in silico* suite OPERA (OPEn structure–activity/property Relationship App) (Mansouri et al., 2018). OPERA is available as a standalone tool, but also as a Konstanz Information Miner (KNIME) workflow, a plugin for the OECD QSAR Toolbox and as part of NICEATM's Integrated Chemical Environment (ICE) platform.

Adequacy of the method for the intended purpose

Where applicable, the CATMoS suite produces both, a predicted ATE value in the form of an LD50 estimate and a prediction of the appropriate GHS/CLP Acute Tox. sub-category. Therefore, it can be concluded that, in principle, CATMoS provides the acute toxicity outputs typically required by EU chemicals legislation.

Reliability and accuracy of the method for the intended purpose

The predictive performance of the CATMoS tool has been broadly investigated by Mansouri et al. (2021) based on a dataset of > 11 000 chemicals. Another evaluation with a specific focus on pesticides was recently published by Bishop et al. (2024).

Whether or not the predictive performance of the method is considered sufficient to replace the current *in vivo* test methods under EU chemicals legislation remains to be ascertained based on an independent assessment and the generic performance benchmarks yet to be developed.

In addition, an analysis of the coverage of chemical space should be performed in order to describe the AD of the method and to identify sectors of chemical space, if any, where the predictivity of the method might not fulfil regulatory expectations. To some degree, such investigations have already been performed in the two publications cited above as well as in a recent publication by industry (Weyrich et al., 2025).

Implementation

Based on the considerations above, it is proposed to implement this strategy in two phases.

Preparatory phase

A small informal working group including experts from ECHA, EFSA, NICEATM, BfR, industry and the European Commission has been set up in January 2024 to perform the following tasks:

- Investigate the composition and quality training and test data sets which have been used to pre-validate CATMoS
- Broaden the test data set with additional curated data from regulatory processes which were not known to the developers of CATMoS (e.g. from ECHA's REACH database and EFSA's openfoodtox database) (Dorne et al., 2021; Patlewicz et al., 2024).
- propose suitable predictivity benchmarks for the prediction of ATE values/ranges, LD50 values, and acute toxicity categories based on the performance of the existing *in vivo* methods and to
- analyse adequacy and reliability of the CATMoS model using the OECD QAF and assess its predictivity against the proposed benchmarks, incl. an estimate of the percentage of substances (within the chemical space relevant for REACH) for which the CATMoS model would provide reliable predictions (according to the benchmarks defined above), taking into account the pre-validation work done already in Mansouri et al. (2018) and Bishop et al. (2024).

Preliminary results

ECHA has started an independent evaluation of AOT data similar to the work described by Karmaus et al. (2022), with the objective of establishing an AOT reference data set for LD50 values based on data from REACH registrations, quantification of the variability between these values and performing a preliminary analysis of performance of the CATMoS model suite. In this context, also general technical questions around building a reference data set, such as handling multiple data points per reference substance, are addressed with a view to extract learnings also relevant for further endpoints later in the roadmap work. Preliminary results indicate that acceptable (from a technical point of view, i.e. not necessarily correct) CATMoS predictions can be obtained for the bulk of REACH registration data, raising hope that, acceptable predictivity assumed, CATMoS may be suitable for replacing AOT *in vivo* testing for a significant part of REACH-relevant chemical space.

Future work will then integrate the learnings from this exercise with that from other recent publications such as Bishop et al. (2024) or Weyrich et al. (2025). On these grounds, a first proposal for benchmark criteria will be developed which can then be carried over to the implementation phase.

Implementation phase

Once the preparatory work on performance benchmarks and the performance assessment of the CATMoS suite will be completed, which is expected to happen in the course of 2025, the process of regulatory implementation should be initiated. The table below provides a non-exhaustive list of ideas for actions that should be taken in this context, along with proposed main actors as well as third parties who should be consulted in the process.

Action	**Main actors**	**To consult with**
Technical expert group to continue work on performance benchmarks and performance assessment.	ECHA, NICEATM, BfR, EPAA, EFSA, further parties have indicated their interest to join	European Partnership for the Assessment of Risk s from Chemicals (PARC) It is anticipated that after a first proposal for benchmarks and a first assessment of CATMoS' performance become available, these preliminary results will be discussed

Action	Main actors	To consult with
		with the OECD QSAR expert group (and, possibly, further OECD expert groups).
Confirm the interpretation and analysis regarding the CLP and REACH legal texts and guidance as presented above with experts for classification & labelling.	European Commission and ECHA	Meeting of the Competent Authorities for REACH and CLP (CARACAL), ECHA's Risk Assessment (RAC) and Member State Committees (MSC)
Perform analogous analyses for all other legislations in scope of the COM roadmap.	COM and responsible European agencies (e.g. ECHA for biocidal products, EFSA for food-related regulations,[63] etc.	Relevant risk assessment and risk management bodies or committees, e.g. SCCS for cosmetics[64], EFSA Scientific Committee for food-related legislation etc., ECHA's Biocidal Products Committee etc., possibly also the OSOA[65] working group.
Consult with stakeholders.	To be determined based on the organisational structure that will be established for	Various stakeholder representatives: academia, industry, NGOs and policymakers

Once this initial implementation phase has been concluded and the general acceptability of *in silico* methods as well as the availability of at least one suitable specific method has been ascertained, the above-mentioned minor updates of the REACH and CLP guidance documents, as well as to other legislation and guidance in the scope of the COM roadmap, where necessary, should be initiated to promote the use of *in silico* methods for AOT assessment as soon as possible. Possible ways to already promote use of *in silico* methods even while the official guidance has not yet been updated should be explored.

In parallel, the necessary steps to initiate updates to the GHS, CLP and as well as other pieces of legislation in the scope of the roadmap in order to implement *in silico* methods as the default approach could be explored. Contact with regulators and stakeholders, including relevant working groups (e.g. at UN GHS level, the Informal Working Group on Non-Animal Test Methods, IWG NATM) will be made at an early time-point to receive feedback and buy-in.

Outlook

For the moment, the scope of this activity encompasses AOT assessment of substances as well as mixtures, following a component-based approach[66]. It remains to be explored to what extent also acute toxicity via the dermal and inhalation routes can be covered with the available *in silico* tools and whether this can already be included in the present activity or remains to be addressed separately.

Further future challenges include handling substances with multiple or inconsistent LD50 values, extending methods to acute dermal and inhalation toxicity, and ensuring coverage of out-of-domain chemicals, including inorganics. Broader integration of additional models beyond CATMoS is also planned.

[64] Although the Cosmetic Products Regulation (EC) 1223/2009 (CPR) prohibits new in vivo testing in vertebrates, still further discussions about the acceptability of *in silico* methods in general, or of specific *in silico* tools in particular, might be required.
[65] One Substance, One Assessment
[66] Component-based approaches aim at determining mixture properties based on the properties of the mixture's individual constituents, as opposed to a whole mixture approach, which can be followed with the *in vivo* test, but is not compatible with *in silico* methods.

References

Bishop, P. L., Mansouri, K., Eckel, W. P., Lowit, M. B., Allen, D., Blankinship, A., Lowit, A. B., Harwood, D. E., Johnson, T., & Kleinstreuer, N. C. (2024). Evaluation of in silico model predictions for mammalian acute oral toxicity and regulatory application in pesticide hazard and risk assessment. Regulatory Toxicology and Pharmacology, 149, 105614. https://doi.org/10.1016/j.yrtph.2024.105614

Dorne JLCM, Richardson J, Livaniou A, Carnesecchi E, Ceriani L, Baldin R, Kovarich S, Pavan M, Saouter E, Biganzoli F, Pasinato L, Zare Jeddi M, Robinson TP, Kass GEN, Liem AKD, Toropov AA, Toropova AP, Yang C, Tarkhov A, Georgiadis N, Di Nicola MR, Mostrag A, Verhagen H, Roncaglioni A, Benfenati E, Bassan A. (2021). EFSA's OpenFoodTox: An open source toxicological database on chemicals in food and feed and its future developments. Environ Int. 146:106293. doi: 10.1016/j.envint.2020.106293.

Karmaus, A. L., Mansouri, K., To, K. T., Blake, B., Fitzpatrick, J., Strickland, J., Patlewicz, G., Allen, D., Casey, W., & Kleinstreuer, N. (2022). Evaluation of Variability Across Rat Acute Oral Systemic Toxicity Studies. Toxicological Sciences, 188(1), 34-47. https://doi.org/10.1093/toxsci/kfac042

Mansouri, K., Karmaus, A. L., Fitzpatrick, J., Patlewicz, G., Pradeep, P., Alberga, D., Alepee, N., Allen, T. E. H., Allen, D., Alves, V. M., Andrade, C. H., Auernhammer, T. R., Ballabio, D., Bell, S., Benfenati, E., Bhattacharya, S., Bastos, J. V., Boyd, S., Brown, J. B., Kleinstreuer, N. C. (2021). CATMoS: Collaborative Acute Toxicity Modeling Suite. Environ Health Perspect, 129(4), 47013. https://doi.org/10.1289/ehp8495

Patlewicz G, Karamertzanis P, Friedman KP, Sannicola M, Shah I.(2024) A systematic analysis of read-across within REACH registration dossiers. Comput Toxicol. 30:1-15. doi: 10.1016/j.comtox.2024.100304.

Weyrich, A., Peter, N., Watzek, N., Michael, S., Lauer, B., Zwinkmann, J., & Teubner, W. (2025). Can acute oral in vivo toxicity testing for EU REACH be fully replaced by a QSAR method? Evaluation of the CATMoS model using chemical industry data. Regulatory Toxicology and Pharmacology, 162, 105861. https://doi.org/https://doi.org/10.1016/j.yrtph.2025.105861

4.2. Genotoxicity

Background

Genotoxicity is considered an important toxicological effect in the safety evaluation of chemicals and pharmaceuticals. It encompasses three endpoints: gene mutation, clastogenicity (structural chromosomal aberrations), and aneugenicity (chromosome loss or gain). Please see the 'Definitions' section below for more detailed descriptions. Assessment of genotoxicity for regulatory purposes usually involves a tiered approach, starting with a standard battery of *in vitro* mutagenicity tests, followed by *in vivo* testing on the same mutagenicity endpoint in case of a positive *in vitro* test result (Cimino, 2006; Dearfield et al., 2011; Eastmond et al., 2009). Generally, the conclusion from this approach is limited to hazard identification: it relies on a yes/no binary decision. Characterisation of the genotoxic hazard using quantitative analyses has been demonstrated to be feasible and possibly is of added value for risk assessment of chemicals and pharmaceuticals (Beal et al., 2023; Chepelev et al., 2023; Luijten et al., 2020; Nicolette et al., 2021; White et al., 2020); however, this is not yet a standard requirement in the regulations currently in use, as several open questions and uncertainties remain to be addressed (Sachse et al., in preparation).

The current standard battery of *in vitro* genotoxicity tests is rather consistent across different sectors and different regulatory jurisdictions. Commonly it includes a bacterial reverse gene mutation test, a mammalian cell gene mutation test, and/or a mammalian cell chromosomal aberration test (European Commission, 2008, 2009, 2013; ICH, 2008) or a bacterial reverse gene mutation test and a mammalian cell micronucleus test (EFSA Scientific Committee, 2011; SCCS, 2023). The performance of the battery for detecting genotoxic potential has been evaluated for a large set of chemicals that are known to be carcinogenic and/or to induce genotoxic effects in rodents (Kirkland et al., 2011). The results of this analysis indicated that a combination of the bacterial reverse gene mutation test, i.e. the Ames test (OECD Test Guideline (TG) 471), and the *in vitro* mammalian cell micronucleus test (MNvit; OECD TG 487), would be sufficient to reliably predict genotoxic potential of chemicals. The MNvit was preferred over the *in vitro* mammalian chromosome aberration test (OECD TG 473), because the MNvit is capable of detecting both clastogenicity and aneugenicity (while the *in vitro* mammalian chromosome aberration test is primarily suited for detecting clastogenicity).

In the context of the European Commission's roadmap for phasing out animal testing for chemical safety assessments, a subgroup of the Human Health working group experts convened and discussed how *in vivo* testing for genetic toxicity could be minimised, whether classification and hazard or risk assessment would be possible based on *in vitro* outcomes only, and whether there is potential for harmonisation of testing requirements across different legislations in scope of the roadmap. The authors explicitly chose to keep the discussions at a relatively high level of abstraction. The reason for doing so is that genotoxicity is a very severe health effect. Many, if not all, of the proposed actions require detailed knowledge and extensive discussions, often complemented with in-depth analyses of large datasets, to ensure that modified or entirely new approaches for addressing genotoxicity still safeguard public health. Below, five different areas for potential changes are described with a brief summary of the current situation and the main opportunities for reducing animal testing, plus proposed actions for the short-, medium- and long-term.

EPAA Team of experts

The EPAA is currently establishing a new team of experts in genotoxicity that will extensively discuss the areas for potential changes and proposed actions, further define an implementation plan, set out clear goals to deliver non-animal-based approach across sectors, as well as define any change management requirements and responsibilities. The team could coordinate some activities, bring the required knowledge together and support driving forward the roadmap goals by developing a consensus approach to genotoxicity assessment, taking into account related initiatives (see examples in Table 6), similar to the carcinogenicity team (see chapter on carcinogenicity). The team will include representatives of industries

across different sectors, EU agencies (EFSA, ECHA, EMA), the European Commission, Member States and academia.

Table 6. EU and global activities and projects on genotoxicity relevant for the roadmap

Ongoing revision of the GHS criteria for classification in the hazard class germ cell mutagenicity led by the European Union.
OECD: Several projects focused on new test methods, which are at different stages of validation, such as ToxTracker (OECD Project 4.125), comet and micronucleus assays in reconstructed skin models (OECD Project 4.139), **γH2AX/pH3 (OECD Project 4.168), *in vitro* alkaline comet assay (OECD Project 4.190), micronucleus test for nano,** error-corrected next-generation sequencing (OECD Project 4.175)
EFSA: current update of its guidance on genotoxicity assessment applicable to food and feed safety.
PARC: IATA for genotoxicity; IATA for carcinogenicity
HESI Genetic Toxicology Technical Committee: improving the scientific basis of the interpretation of results from genetic toxicology tests for regulatory purposes and promoting and monitoring the development of innovative tests and testing strategies for assessment of genotoxic potential
Health Canada: GeneTox21 platform including high-throughput transcriptomic (TGx-DDI) and flow cytometry-based assays (flow cytometry-based micronucleus (MN) test, multiplexed flow cytometry-based reporter assay)

Reducing *in vivo* testing through changing GHS classification criteria

Under the United Nations Globally Harmonized System for classification and labelling (GHS), classification of chemicals for germ cell mutagenicity is based on their potential to induce heritable mutations in germ cells, within two main categories: Category 1 including known (1A) or presumed (1B) human germ cell mutagens and Category 2 covering chemicals suspected of causing such effects (GHS, 2025). Currently, classification in Category 1A is based on human evidence from epidemiological data, while Category 1B is based on information from animal testing or effects in human germ cells. *In vitro* alternatives for assessing germ cell mutagenicity do not currently exist.

To avoid follow-up *in vivo* germ cell testing after positive results from *in vivo* genotoxicity testing in somatic cells, the Netherlands submitted a proposal to GHS in 2023 to classify also *in vivo* somatic cell mutagens as Category 1. The main rationale for the proposal is that somatic cell mutagens are not appropriately risk managed. Classification of substances as Category 2 for carcinogenicity or reproductive toxicity means that these substances are suspected to cause adverse health effects due to insufficient evidence. Most of the substances suspected to cause germ cell mutagenicity and classified as Category 2 are, however, proven to be somatic cell mutagens. They may be able to reach the gonads and potentially cause germ cell mutagenicity (even if not proven). Somatic cell mutagenicity can therefore be considered a critical health hazard in itself due to its high correlation with carcinogenicity (90% Annex VI to CLP substances[67]). Therefore, the possibility to waive *in vivo* studies, such as the two-year rodent cancer bioassay, based on evidence for somatic cell mutagenicity would lead to a considerable reduction of animal

[67] Annex VI to CLP, https://echa.europa.eu/information-on-chemicals/annex-vi-to-clp

testing. The discussion on the extension of the hazard class germ cell mutagenicity to mutagenicity (covering both germ cell and somatic cell mutagens) is proposed to be postponed until completion of the ongoing revision of the germ cell mutagenicity chapter in the GHS. The ongoing revision of the GHS chapter also includes exploring possibilities to use somatic cell mutagenicity data in combination with indirect evidence of interaction with germ cells to classify chemicals as Category 1B. In addition, non-testing methods such as read-across and (Q)SARs are introduced in the chapter on germ cell mutagenicity in the GHS as suitable evidence for classification.

Proposed Actions

Short-term:

- Finalise the revision of the GHS chapter on germ cell mutagenicity: reach consensus on incorporating information from *in vitro* and non-animal test methods wherever possible (*Actors: GHS Informal Working Group*).

Medium -term:

- Restart the discussion at the GHS to either extend the current hazard class on germ cell mutagenicity to mutagenicity or to introduce an additional class for mutagenicity (*Actors: GHS Subcommittee*).
- Explore options at EU level to increase risk management measures for Category 2 germ cell mutagens, in case no consensus is achieved to change classification criteria or to introduce a hazard class for mutagenicity, i.e. trigger CMR Category 1 risk management measures for Category 2 germ cell mutagenicity in EU chemical legislation. This would lead to the same desired level of protection while omitting the need for *in vivo* germ cell genotoxicity testing (*Actors: DG ENV, DG GROW, ECHA*).

Reducing or replacing *in vivo* testing through improved in vitro methods

Various *in vitro* test methods for the assessment of genotoxicity are available, including tests that are described as OECD TGs and that are commonly used for regulatory purposes. These standard *in vitro* genotoxicity tests are considered sensitive, but their specificity is relatively low: *in vitro* tests for genotoxicity are optimised to identify the presence of a genotoxicity hazard. While this is a desired outcome from a regulatory perspective, it also leads to a relatively high false positive rate (although some recent dosing advice changes applied to the OECD TGs has substantially reduced the potential for false positives). A positive *in vitro* result may, depending on the regulatory jurisdiction, trigger *in vivo* follow-up testing, which is sometimes unnecessary in case this *in vitro* result is indeed a false positive (for example under REACH). Therefore, identifying the false positive *in vitro* results early on would avoid following up with *in vivo* studies that would yield negative test results. As such, improvement of the composition of the standard *in vitro* battery and/or providing *in vitro* alternatives to be used as follow-up to confirm (or deny) an initial positive *in vitro* mutagenicity test result would contribute to a reduction in animal testing for genotoxicity assessment. This requires the development and validation/qualification of more human-relevant *in vitro* test methods (including but not limited to ongoing work at OECD, such as ToxTracker assay, MultiFlow assay, DNA sequencing, biomarker-based tests) addressing all genotoxic endpoints. Key features of more human-relevant *in vitro* test methods are adequate metabolic capacity as well as high specificity (besides high sensitivity). Also, different test methods should cover different cell types.

Improvement of the standard battery for genotoxicity assessment, and thus the development of *in vitro* follow-up tests, should go beyond simply enhancing test specificity. Currently, the standard battery consists of tests that offer limited mechanistic insight and are relatively low in throughput. To progress towards a paradigm where *in vitro* testing only is sufficient, new methods must deliver mechanistic data that reveal disruption in key biological events leading to genotoxic or mutagenic effects. In combination, these methods should encompass different modes of action (MoA) relevant for genotoxicity and more accurately replicate

the *in vivo* exposure environment, including factors such as metabolic activation, DNA repair capacity, and oxidative stress responses.

Phasing out animal tests for genotoxicity assessment requires a deeper understanding of toxicokinetic (ADME) parameters and effective strategies for generating and applying this information to improve the predictive power of *in vitro* methods. While progress has been made, current advancements remain insufficient to support the full implementation of an exclusively *in vitro* testing approach. More critically, transitioning to such a paradigm demands careful consideration of the intended protection goals. This includes evaluating where and how quantitative analysis of genotoxic effects can add value. Ultimately, a robust chemical risk assessment framework should integrate mechanistic insights, potency data and toxicokinetic profiles.

Proposed Actions

Short-term:

- Complete validation according to OECD GD 34 of alternative approaches for genotoxicity that are already in the pipeline.
- (Further) Develop more human-relevant genotoxicity *in vitro* test methods using different cell types. Key features of such more human-relevant *in vitro* test methods are adequate metabolic capacity as well as high specificity (*Possible actors: PARC, test method developers*).
- Discuss with different stakeholders involved the desired level of protection and thus the adequate level of prediction for innovative test methods for genotoxicity assessment (see also chapter on level of protection in this SWD).
- Develop case studies aimed at exploring how to best use toxicokinetic information to refine the prediction of *in vitro* test methods for genotoxicity assessment (*Possible actors: PARC, EPAA, HESI GTTC*).
- Introduce potency in the evaluation of *in vitro* tests for genotoxicity to allow distinguishing between weak/moderate and strong mutagens; case studies will be needed to demonstrate where and how this could be of added value.
- Explore (sand box) the contexts of use in which a threshold for genotoxicity can be assigned and/or quantitative analyses can be used to derive a point of departure for risk assessment.

Medium-term:

- Continue to develop advanced test systems with higher cellular complexity such as 3D models for liver, lung, gastrointestinal tract.
- Achieve qualification and/or validation of promising more human-relevant, metabolically competent genotoxicity *in vitro* test methods using different cell types.
- Discuss regulatory uptake of a threshold and/or point of departure for genotoxicity in chemical risk assessment.

Reducing *in vivo* testing through application of a weight of evidence assessment in certain cases

Negative results obtained from an adequate evaluation with the standard *in vitro* battery for genotoxicity can be considered sufficient to conclude that the substance is non-mutagenic in some but not all regulatory areas. While there is a need to modernise the standard battery to implement more relevant *in vitro* test systems with adequate metabolic capacity and improved specificity as outlined above, the high sensitivity of the current standard *in vitro* battery can be retained, while considering the possibility of false negatives (due to e.g. limited cellular uptake, *in vitro*-specific metabolism, low solubility, high volatility). If positive results are obtained in the standard *in vitro* battery, *in vivo* follow-up testing may be required. Before

proceeding directly to *in vivo* testing, a weight of evidence (WoE) assessment may be performed in certain scenarios, considering both existing and newly generated non-animal information to evaluate the possibility of avoiding *in vivo* confirmation. When routinely integrated into the follow-up testing and assessment strategy, the WoE approach may, in some cases, allow the conclusion that a substance is (not) mutagenic without the need to conduct *in vivo* genotoxicity tests.

The WoE approach should address whether the positive or negative *in vitro* result is predictive of mutagenic effects in humans. Relevant information to be considered in the WoE approach may be obtained from existing data (e.g. testing data on physico-chemical, toxicokinetic, and toxicodynamic properties; non-testing data such as (Q)SAR, kinetic simulations, and read-across) or by generating new data using a Non-Animal Method-based strategy. The possibility of false/misleading positives should be addressed considering factors known to influence the specificity of *in vitro* assays (e.g. *in vitro* specific metabolism, cytotoxicity, altered testing conditions like pH, osmolality, precipitates, etc.). Information on the expected absorption, distribution, metabolism, and excretion (ADME) of the substance in humans is required. Such toxicokinetic information is needed, for example, to evaluate systemic availability and identify potential target tissues (initial site of contact, such as skin, respiratory tract, upper gastrointestinal tract) versus systemic targets. Information on the metabolism of the substance should be gathered to decide whether the *in vitro* system adequately covers the relevant metabolic pathways. Toxicokinetic information may be obtained from available *in vivo* ADME data or may be generated *in vitro* or through kinetic modelling. Existing information on toxicological endpoints other than genotoxicity may be used to better understand the toxicokinetic (e.g. systemic availability) and toxicodynamic (e.g. target tissue/organ) properties. Non-testing data obtained through e.g. read-across or (Q)SAR applications may provide information on the mutagenic mechanism.

To follow-up and confirm the mutagenic effects (or lack thereof) observed with the standard *in vitro* battery, higher-tier 3D tissue models can be used, which are generally considered to be a more biologically relevant *in vivo*-like test system due to their higher cellular complexity. Advanced 3D *in vitro* models are already recommended in the cosmetics sector, and the development of OECD TGs for 3D Skin models is underway (OECD Project 4.139: 3D reconstructed skin micronucleus test and comet assay). Other 3D models (e.g. liver, lung) are expected to become available in the near future. The mutagenic mode of action (MoA) is another important line of evidence for assessing relevance to humans. While several alternatives to animal approaches (e.g. omics methods, error-corrected sequencing (ECS, OECD Project 4.175) or reporter assays (e.g. Toxtracker, OECD Project 4.125) have been developed to elucidate the MoA, ongoing efforts are underway to develop Adverse Outcome Pathways (AOPs) and AOP networks to provide a framework for organising mechanistic information. Methods employing ECS in cell lines of different origin (rodent vs. human cell lines) may be used to inform cross-species extrapolation. The Notes of Guidance published by the SCCS (2023) already provides valuable advice.

If the WoE assessment does not confirm the initial positive or negative result, further investigation, including *in vivo* testing, may be needed. Once validated, non-animal testing methods that reliably predict *in vivo* effects with high sensitivity and specificity; may be sufficient to overrule positive results from the initial *in vitro* test battery. Such methods could support a conclusion that the substance lacks relevant mutagenic properties in humans.

Proposed Actions

Short-term:

- Perform evaluation of available non-animal strategies, e.g. Cosmetics (*Possible actor: EPAA*).
- Conduct case studies for different contexts of use to identify scenarios in which an *in vitro*-only testing strategy may provide sufficient WoE to conclude on genotoxic potential (*Possible actors: PARC, EPAA*).

- Continue to develop and improve toxicokinetic models, e.g. to better predict systemic availability and tissue/organ distribution (*Possible actors: academia, industry*).
- Continue to develop AOPs and AOP networks for genotoxicity (*Possible actors: PARC, academia, industry*).
- Discuss with different stakeholders involved the desired level of protection and thus the adequate level of prediction for innovative test methods for genotoxicity assessment (see also chapter on level of protection in this SWD).

Medium-term:

- Develop a decision tree for the WoE assessment to provide guidance on when additional Non-Animal Method-based information is sufficient to avoid subsequent *in vivo* testing (e.g. updating Agencies Guidances on genotoxicity). The decision tree for a testing strategy for cosmetic ingredients (SCCS, 2023) can serve as a starting point for further development and adoption in other regulatory areas. Non-animal approaches for ADME should where relevant be included.
- Develop case studies integrating (mechanistic) *in vitro* test methods for toxicokinetics and genotoxicity into a defined approach / WoE approach (*Possible actors: PARC, academia, industry*).
- Develop an OECD Test Guideline for a defined approach / WoE approach for genotoxicity assessment, including data interpretation procedure and guidance.
- Update respective EU-regulations to allow for hazard identification and classification based on *in vitro* data.

Long-term:

- Define the information requirements for a WoE assessment that allows to overrule a positive result in the standard *in vitro* battery without follow-up *in vivo* testing.
- Modernize the standard *in vitro* battery for genotoxicity assessment.

Updating/improving current testing strategies

Current genotoxicity testing strategies are described in the revised REACH endpoint specific guidance for mutagenicity (IR&CSA Chapter R.7a part R.7.7), in the ICH guideline on genotoxicity testing (2008), the EFSA Scientific Opinion on genotoxicity testing strategies (2011), and the SCCS Notes of Guidance (2023). Further, work is ongoing under PARC for the development of an AOP-based IATA for genotoxicity. Actions listed in this document are expected to lead to changes of the above-mentioned testing strategy documents. Also, the evaluation of harmonisation across legislations as laid out in this document may inform changes to the respective guidance documents.

A stepping stone on the way to full replacement of *in vivo* testing for genotoxicity is reduction of testing under the current regulatory paradigms. For example, in a somatic cell gene mutation study in transgenic rodents, the collection and storage of germ cells for potential further analysis – should the study yield a positive result – can eliminate the need for a separate germ cell mutagenicity study.

Integration of genotoxicity endpoints into regulatory mandated *in vivo* studies offers another possibility to reduce animal numbers and animal testing in the short- to medium-term. The combination of the *in vivo* comet assay with the *in vivo* micronucleus assay has been described in the respective OECD TGs 489 and 474 to reduce animal numbers as compared to running the two methods independently (Hamada et al., 2001; Pfuhler et al., 2009; Bowen et al., 2011). The Pig-a gene mutation assay can also be combined with those assays as described in OECD TG 470. Further, the integration of genotoxicity endpoints into repeated-dose toxicity studies has been evaluated by an IWGT (International Workshop on Genotoxicity Testing) working group (Pfuhler et al., 2009; Rothfuss et al., 2011; EFSA 2011) and is mentioned in ECHA guidance in the revised IR&CSA Chapter 7.7. The previously issued guidance on combined protocols, with regards to treatment schedules, exposure duration and dose-level setting should be critically reviewed, as well as any published examples that support the guidance. The development and publication of new case studies

can provide robust evidence supporting the feasibility of meeting regulatory requirements with combined studies.

Proposed Actions

Short-term:

- Identify needs to further work on integrated *in vivo* studies to reduce animal numbers while ensuring scientific robustness and regulatory acceptance of studies.
- Conduct case studies illustrating combined study protocols (*Possible actors: PARC, HESI GTTC, Member States*).
- Integrate actions and developments into an aligned testing strategy.
- Update guidance documents for genotoxicity testing strategies and integrating *in vivo* study designs (*Actors: DGs, ECHA, EFSA, SCCS*).

Harmonisation of in vitro and *in vivo* testing requirements under different legislations

Genotoxicity assessment is an essential component of the safety assessment of all types of substances across sectors. Although the assessments have many elements in common (typically starting with testing in an *in vitro* battery followed up by *in vivo* testing in case of positive *in vitro* results) they also present several differences, which are due to various reasons (type of exposure, perceptions, history). Therefore, this is an area where there is room for harmonisation, which would allow for a common evaluation of substances under different legislative frameworks in line with the one-substance-one-assessment concept (OSOA).

For instance, two evident differences in approaches include: a) the number of *in vitro* tests required in the basic battery under different regulated sectors and b) whether or not *in vivo* studies are required in case of negative *in vitro* battery results (e.g. they are mandatory for pharmaceuticals and plant protection products, but not for industrial chemicals).

A suitable group to addressing such harmonisation issues could be the newly established EPAA working group on genotoxicity, especially due to its cross-sector composition and inclusion of EU Agencies representatives.

Proposed Actions

Short-term:

- Identify the main differences in approaches across sectors, understand the related reasons for such differences and evaluate the justifications for such differences (*Possible actors: EPAA, PARC*).

Short- to medium-term:

- Analyse the identified varying elements of assessment strategies across sectors and conclude whether these can be harmonised without reducing the current level of protection, else define what other lines of evidence will provide confident information (*Possible actor: EPAA*).

Medium-term:

- Engage with EU and Member State regulators across sectors to amend regulatory requirements towards a harmonised approach and the implementation of the OSOA concept, where possible (*Possible actors: Member States, COM, Agencies*).

Definitions

The term *mutation* applies both to permanent heritable and non-heritable genetic changes that may be manifested at the phenotypic level and to the underlying DNA modifications (including, for example, specific base pair changes and chromosomal translocations).

The term *gene mutation* refers to permanent changes in the base sequences of a certain gene. The term *clastogenicity* refers to structural chromosome aberrations. A clastogen can cause the loss or rearrangements of chromosome segments. *Aneugenicity* (aneuploidy induction) refers to a change (loss or gain) in chromosome number in cells (numerical chromosome aberrations).

The term *genotoxicity* refers to permanent (i.e. *mutation*) as well as reversible changes in genetic information. Both processes could be mediated by *primary DNA damage*. Primary DNA damage may be reversed by DNA repair or other cellular processes, and, thus, may or may not result in permanent alterations in the structure or information content of the genetic material in a surviving cell or its progeny. Genotoxicity thus encompasses mutations, including gene mutation, clastogenicity and aneugenicity, as well as primary DNA damage.

Primary DNA damage can be identified using indicator tests to assess effects such as, DNA strand breaks and DNA adduct formation. Genotoxicity tests include mutagenicity tests and indicator tests.

References

Beal MA, Audebert M, Barton-Maclaren T, Battaion H, Bemis JC, Cao X, Chen C, Dertinger SD, Froetschl R, Guo X, et al. 2023. Quantitative in vitro to in vivo extrapolation of genotoxicity data provides protective estimates of in vivo dose. Environmental and molecular mutagenesis 64: 105-122.

Bowen DE, Whitwell JH, Lillford L, Henderson D, Kidd D, Mc Garry S, Pearce G, Beevers C and Kirkland DJ (2011) Work conducted at Covance Laboratories Ltd., Harrogate. Evaluation of a multi-endpoint assay in rats, combining the bone-marrow micronucleus test, the Comet assay and the flow-cytometric peripheral blood micronucleus test. Mutat Res 722: 7-19.

Chepelev N, Long AS, Beal M, Barton-Maclaren T, Johnson G, Dearfield KL, Roberts DJ, van Benthem J, White P. 2023. Establishing a quantitative framework for regulatory interpretation of genetic toxicity dose-response data: Margin of exposure case study of 48 compounds with both in vivo mutagenicity and carcinogenicity dose-response data. Environmental and molecular mutagenesis 64: 4-15.

Cimino MC. 2006. Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes. Environmental and molecular mutagenesis 47: 362-390.

Dearfield KL, Thybaud V, Cimino MC, Custer L, Czich A, Harvey JS, Hester S, Kim JH, Kirkland D, Levy DD, et al. 2011. Follow-up actions from positive results of in vitro genetic toxicity testing. Environmental and molecular mutagenesis 52: 177-204.

Eastmond DA, Hartwig A, Anderson D, Anwar WA, Cimino MC, Dobrev I, Douglas GR, Nohmi T, Phillips DH, Vickers C. 2009. Mutagenicity testing for chemical risk assessment: Update of the who/ipcs harmonized scheme. Mutagenesis 24: 341-349.

European Commission. 2008. Council regulation (EC) no 440/2008 of 30 May 2008 laying down test methods pursuant to regulation (EC) no 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). Official Journal of the European Union L 142.

European Commission. 2009. Regulation (EC) no 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council directives 79/117/EEC and 91/414/EEC. Official Journal of the European Union L 309/1.

European Commission. 2009. Regulation (EC) no 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. Official Journal of the European Union L 342.

European Commission. 2013. Commission regulation (EU) no 284/2013 of 1 March 2013 setting out the data requirements for plant protection products, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. Official Journal of the European Union L 93/85.

European Commission. 2025. Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. Official Journal of the European Union. Available at: http://data.europa.eu/eli/reg/2008/1272/2022-12-17

European Food Safety Authority. 2011. Scientific opinion on genotoxicity testing strategies applicable to food and feed safety assessment. EFSA Scientific Committee. EFSA Journal 2011;9(9):2379.

Globally Harmonized System of Classification and Labelling of Chemicals (Rev.11) (2025) ST/SG/AC.10/30/Rev.11. https://unece.org/sites/default/files/2025-09/GHS%20Rev11e.pdf

Hamada S, Sutou S, Morita T, Wakata A, Asanami S, Hosoya S, Ozawa S, Kondo K, Nakajima M, Shimada H, Osawa K, Kondo Y, Asano N, Sato S, Tamura H, Yajima N, Marshall R, Moore C, Blakey DH, Schechtman LM, Weaver JL, Torous DK, Proudlock R, Ito S, Namiki C and Hayashi M (2001) Evaluation of the rodent micronucleus assay by a 28-day treatment protocol: Summary of the 13th Collaborative Study by the Collaborative Study Group for the Micronucleus Test (CSGMT)/Environmental Mutagen Society of Japan (JEMS)-Mammalian Mutagenicity Study Group (MMS). Environ Mol Mutagen 37:93-110.

ICH 2008. Guideline S2 (R1) on genotoxicity testing and data interpretation for pharmaceuticals intended for human use.

Luijten M, Ball NS, Dearfield KL, Gollapudi BB, Johnson GE, Madia F, Peel L, Pfuhler S, Settivari RS, Ter Burg W, et al. 2020. Utility of a next generation framework for assessment of genomic damage: A case study using the industrial chemical benzene. Environmental and molecular mutagenesis 61: 94-113.

Kirkland D, Reeve L, Gatehouse D, Vanparys P 2011. A core in vitro genotoxicity battery comprising the Ames test plus the in vitro micronucleus test is sufficient to detect rodent carcinogens and in vivo genotoxins. Mutation Research 721:27-73.

Nicolette J, Luijten M, Sasaki JC, Custer L, Embry M, Froetschl R, Johnson G, Ouedraogo G, Settivari R, Thybaud V, Dearfield KL. 2021. Utility of a next-generation framework for assessment of genomic damage: A case study using the pharmaceutical drug candidate etoposide. Environmental and molecular mutagenesis 62: 512-525.

Pfuhler S, Kirkland D, Kasper P, Hayashi M, Vanparys P, Carmichael P, Dertinger S, Eastmond D, Elhajouji A, Krul C, Rothfuss A, Schoening G, Smith A, Speit G, Thomas C, van Benthem J and Corvi R (2009) Reduction of use of animals in regulatory genotoxicity testing: Identification and implementation opportunities—Report from an ECVAM workshop. Mutat Res 11 680: 31–42.

Rothfuss A, Honma M, Czich A, Aardema MJ, Burlinson B, Galloway S, Hamada S, Kirkland D, Heflich RH, Howe J, Nakajima M, O'Donovan M, Plappert-Helbig U, Priestley C, Recio L, Schuler M, Uno Y and Martus HJ (2011) Improvement of in vivo genotoxicity assessment: combination of acute tests and integration into standard toxicity testing. Mutat Res 723:108-20.

Sachse et al. (2026) Quantitative Evaluation of Genotoxicity Data for Risk Assessment and Regulatory Decision-making: Time for a Paradigm Shift. Environment International 208: 110070.

SCCS (Scientific Committee on Consumer Safety) 2023. SCCS Notes of Guidance for the testing of cosmetic ingredients and their safety evaluation 12th revision. SCCS/1647/22.

White PA, Long AS, Johnson GE. 2020. Quantitative interpretation of genetic toxicity dose-response data for risk assessment and regulatory decision-making: Current status and emerging priorities. Environmental and molecular mutagenesis 61: 66-83.

4.3. Carcinogenicity

Implementing change for carcinogenicity assessments

Summary

Carcinogenicity is considered an important toxicological effect in the safety evaluation of chemicals and pharmaceuticals. The rodent cancer bioassay, currently the standard method for assessing the carcinogenic potential of chemicals, has several limitations. It is time-consuming, requires large numbers of animals across one or two bioassays (rats and mice), has poor reproducibility and has limited ability to predict human carcinogenicity. To address these limitations, new approaches are being developed to modernize carcinogenicity assessment.

A new approach for carcinogenicity assessment presents an excellent case example for regulatory implementation of alternatives to animal approaches for complex endpoints and implications for translation of novel approaches into practical solutions. Multiple initiatives are underway or already in use in a specific area (see ICHS1B guidance for pharmaceuticals) to develop frameworks for a new approach to carcinogenicity assessment, incorporating weight of evidence (WoE)-based methodologies that utilize *in silico*, *in vitro*, and, while still necessary until non-animal approaches are considered robust, short-term *in vivo* (toxicity) tests. In the long term, the intention is to move towards an animal-free approach.

To facilitate the development of a harmonized framework for a WoE-based carcinogenicity assessment, the European Partnership for Alternative Approaches to Animal Testing (EPAA) has established a working group that includes representatives from related initiatives, ensuring synergy and co-creation. The working group aims to identify opportunities for implementing a change for carcinogenicity assessments, with focus on achieving risk management goals that meet regulatory authority standards worldwide within a realistic timeframe, across all sectors.

A review of ongoing initiatives across regulatory areas and jurisdictions revealed substantial commonalities, as they are all being built upon mechanistic approaches to identify potential carcinogenicity concerns earlier in the assessment process, i.e. prior to or instead of proceeding to a bioassay. Moreover, they share a common rationale and justification for using alternative approaches, i.e. the need for a quicker assessment, higher human relevance and a reduction in the number of animals used. This led to the development of an overarching carcinogenicity assessment framework, comprising various components that experts and stakeholders must consider in order to identify current and future information needs that will enable technological advances to keep pace with a changing regulatory paradigm. A harmonized approach, at least in Europe and ideally globally, is essential for successful implementation. However, this will require addressing differences in regional legislation and information requirements, identifying commonalities, and determining what regulatory changes are needed to implement a new approach. Effective communication with expert committees from governmental organizations, stakeholders, and the general public is also crucial.

Taking all this into consideration, a stepwise plan is proposed, outlining short-, medium-, and long-term actions designed to bridge scientific knowledge gaps and implement the necessary measures for a paradigm change. The goal is to inform and advance the roadmap for phasing out animal testing, ensuring a more efficient, innovative, effective and humane approach to carcinogenicity assessment.

History of the approach: why changing carcinogenicity assessment?

The rodent cancer bioassay (OECD Test Guidelines 451/453; OECD, 2018a, 2018b), is currently used worldwide as a standard, default approach for assessing the carcinogenic potential of most chemicals[68] across various sectors. This bioassay (fundamentally unchanged since the 1960's) is time intensive, uses large numbers of animals, and has limited ability to predict relevant human carcinogenic risk. To overcome these shortcomings, new approaches are being developed to facilitate the modernization of carcinogenicity assessment. A carcinogenicity assessment includes the evaluation of both genotoxic and non-genotoxic activity. Acknowledging that genotoxicity evaluation is an integral part of the carcinogenicity assessment, detailed actions for genotoxicity are discussed in a separate chapter of this SWD.

Mapping Global Initiatives and Securing the Learnings

Several initiatives are currently underway to develop a new approach for carcinogenicity assessment, incorporating WoE-based assessments that utilize *in silico*, *in vitro*, short-term *in vivo* tests and/or other available data with the final aim to move towards an animal-free assessment. Bringing together these initiatives is key given the challenges in evaluating carcinogenic potential of chemicals and pharmaceuticals and ensuring a harmonized scientific approach. To foster understanding, the EPAA working group on carcinogenicity listed the efforts in different sectors and jurisdictions (see Table 7 for details). It should be acknowledged that the map might not necessarily be comprehensive. The primary objective of this exercise was to identify areas of commonality, differences, gaps and potential incompatibility, with the ultimate goal of developing a harmonized WoE-based approach for carcinogenicity assessment across industry sectors, including industrial chemicals, agrochemicals, biocidal products, food contact materials, food and feed ingredients, cosmetic ingredients, pharmaceutical products and contaminants.

The global scope of these initiatives is a crucial consideration when striving for synergies and concepts. The numerous efforts undertaken to date have yielded valuable insights that should be preserved and built upon moving forward. An analysis of the various initiatives revealed several key commonalities, particularly their mechanism-based approach, which enables the identification of potential carcinogenicity concerns. This mapping exercise aimed at facilitating the identification and alignment of scientific building blocks, ultimately contributing to the development of a unified approach for carcinogenicity assessment that can be applied consistently across different sectors. While some differences in approach were noted, no significant incompatibilities were identified, suggesting that these initiatives can be aligned and integrated to form a cohesive framework for carcinogenicity assessment[69].

Description of the readiness, applicability domain and validation status of single alternatives to animal approaches that are possibly relevant for carcinogenicity assessments of chemicals and pharmaceuticals is not covered here as that is currently being addressed by the OECD non-genotoxic carcinogenicity Integrated Approach to Testing and Assessment (IATA) expert group (further referred to as OECD NGTxC IATA expert group; Jacobs et al., 2020).

Table 7. EU and global initiatives considered in the mapping effort.

Initiative	Reference

[68] This no longer applies to pharmaceuticals since the addendum to the globally harmonised guidance ICHS13B which establishes a new carcinogenicity assessment framework for pharmaceuticals.

[69] A unified approach to carcinogenicity assessment is considered useful, even though the conclusions from it could differ in the end, as for pharmaceuticals, a benefit-risk assessment must be conducted while for industrial chemicals, no benefits are considered in the risk management decision.

EPAA Mechanistic prediction of carcinogenicity for agrochemicals	doi: 10.1080/10408444.2020.1841732
ReCAAP (Rethinking chronic toxicity and Carcinogenicity Assessment for Agrochemicals Project) framework	OECD (2024), Case Study on the Use of Integrated Approaches for Testing and Assessment for Chronic Toxicity and Carcinogenicity of Agrichemicals with Exemplar Case Studies, OECD Series on Testing and Assessment, No. 402, OECD Publishing, Paris. https://doi.org/10.1016/j.yrtph.2022.105160 https://doi.org/10.3389/ftox.2024.1394361
OECD IATA for Non-Genotoxic Carcinogens	https://doi.org/10.1007/s00204-024-03753-y https://doi.org/10.1007/s00204-020-02784-5
RISK-HUNT3R (RISK assessment of chemicals integrating HUman centric Next generation Testing strategies promoting the 3Rs)	https://www.risk-hunt3r.eu/
PARC (Partnership for the Assessment of Risks from Chemicals)	IATA for genotoxicity; IATA for carcinogenicity https://www.eu-parc.eu/
HESI eSTAR Transcriptomic Data Files for Carcinogenicity	https://doi.org/10.1093/toxsci/kfac041
ICHS1B Testing of Carcinogenicity of Pharmaceuticals - (R1) Addendum - WoE Approach	doi: 10.3389/ftox.2024.1353783 doi/10.1111/bcp.15790
OECD Guidance on Good Practices and Standardization of Sample Collection for Omics Analysis	Accepted by OECD WPHA (June session, to be declassified Q4 2025)

Proposed Framework for Global Carcinogenicity Assessment of Chemicals

The summary above highlights the scientifically agreed importance of a WoE-based approach for carcinogenicity assessment (including the CLP regulation; EC, 2025), clearly providing an opportunity for harmonization. By integrating the key elements from various initiatives, a proposed overarching framework for carcinogenicity assessment has been developed by the EPAA working group. This framework is depicted in Figure 8. The proposed framework might be applied uniformly across sectors and globally, although it is reasonable to expect that details regarding the tiering and decision criteria will vary across sectors, reflecting specific risk management approaches. Furthermore, the framework can be used for different contexts of use, addressing regulatory questions related to both cancer hazard (including the CLP regulation) and risk assessment. To achieve a globally harmonized chemical safety assessment, it is essential to establish consistent methodologies for WoE-based evaluation of carcinogenic potential.

The proposed global framework for carcinogenicity assessment (Figure 8) incorporates a number of tiered elements starting with problem formulation and understanding the external exposure or use scenarios in order to evaluate the exposure-based relevance for carcinogenic risk.

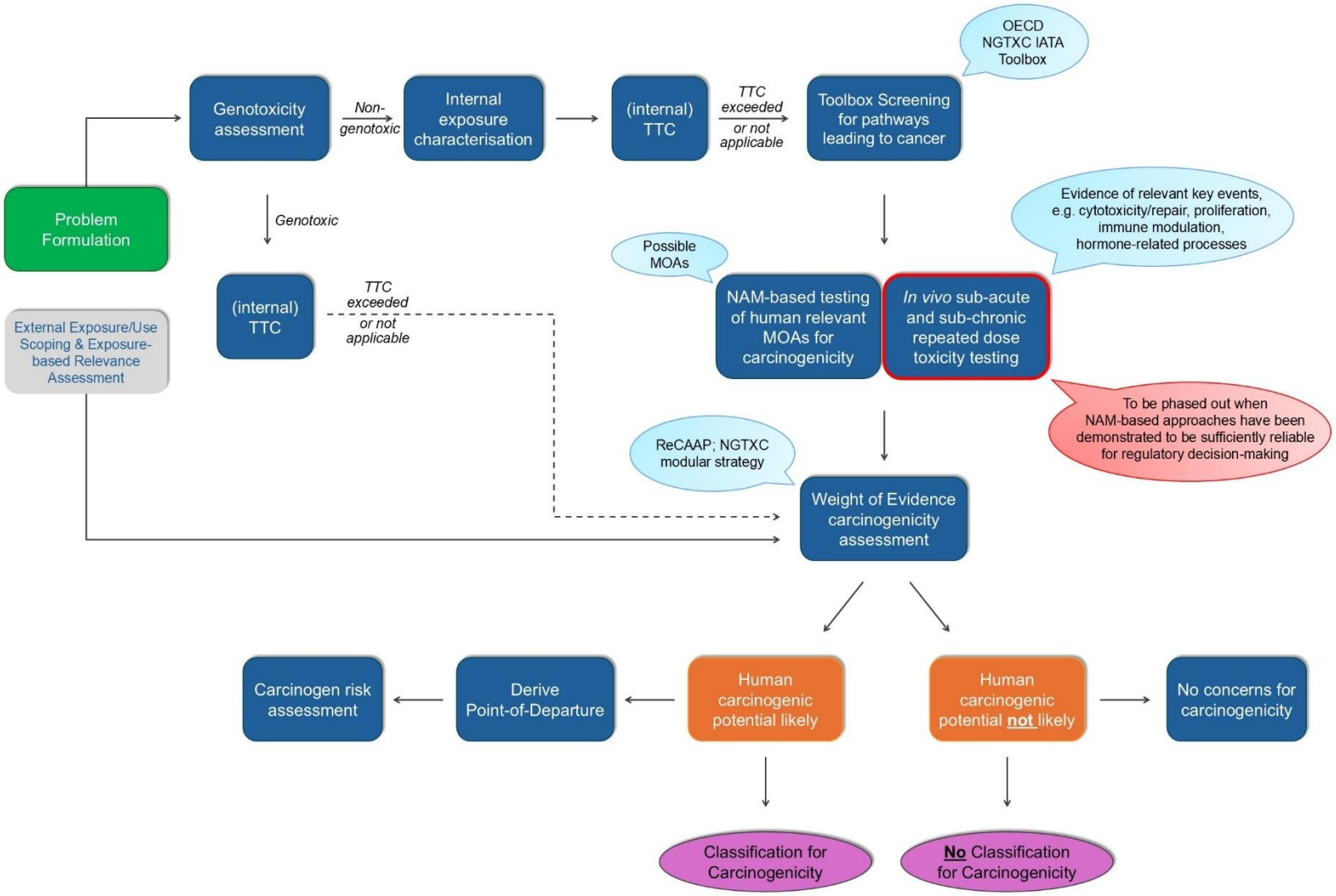
If no exposure potential is identified, or a clear carcinogenic hazard is already known, a justification may be formulated to directly address these risks, and no further data generation would be required. The likelihood of a human carcinogenic potential can be defined and appropriate next steps followed, either to determine classification for labelling or derive a point of departure for risk assessment, as appropriate to the legislative requirements.

If exposure potential is identified or cannot be appropriately characterized, the first step is to conduct a genotoxicity assessment, using Non-Animal Method-based approaches.

If, based on the genotoxicity assessment (see chapter on Genotoxicity), a genotoxicity risk is identified, an (internal) Threshold of Toxicological Concern (iTTC) approach could be used to identify a threshold for risk assessment. If the TTC for potential DNA-reactive mutagens and/or carcinogens is exceeded or the TTC approach is not applicable, a WoE approach could be applied. If the potential Margin of Exposure (MoE) is ≥10,000 (based on the current EFSA Guidance (EFSA, 2025); to be updated for NAM-based approach), risk management may still be feasible under certain legislation. Based on the genotoxic Mode of Action (MoA) identified, classification for mutagenicity would also be determined where required (see chapter on Genotoxicity). For example, it is envisaged that classification for carcinogenicity may not be needed for substances classified as (germ cell) mutagens Category 1 (GHS/CLP), as these would in any case undergo severe risk management measures.

If no genotoxicity risk is identified, a non-genotoxicity carcinogen assessment has to be conducted. First, internal exposure should be characterized to understand systemic risk, and this could be compared with an (intl) TTC for non-genotoxic substances for risk assessment. If the (internal) TTC is exceeded (or this approach is not applicable), evaluation would continue using a screening toolbox to determine potential pathways leading to cancer, based on the OECD IATA for non-genotoxic carcinogens (Jacobs et al., 2020). This toolbox allows NAM-based identification of possible human relevant MoAs for carcinogenicity (Heusinkveld et al., 2020). During the transition phase towards only using non-animal approaches, the toolbox could be supplemented with *in vivo* subacute and/or subchronic repeated dose toxicity testing to provide further evidence for relevant key events. These *in vivo* studies will be phased out when NAM-based approaches have been demonstrated to be sufficiently reliable and relevant for regulatory decision-making. These various data will then be evaluated in a WoE assessment, using e.g. the ReCAAP framework (Hilton et al., 2022), the modular strategy (Louekari and Jacobs, 2024), and the OECD WoE assessment scheme (Figure 9), to define the likelihood of a human carcinogenic potential. These data will then inform classification determinations and derive a point of departure for relevant risk assessments, as appropriate to the legislative requirements.

Global Carcinogenicity Assessment Framework for Chemicals



MoA:

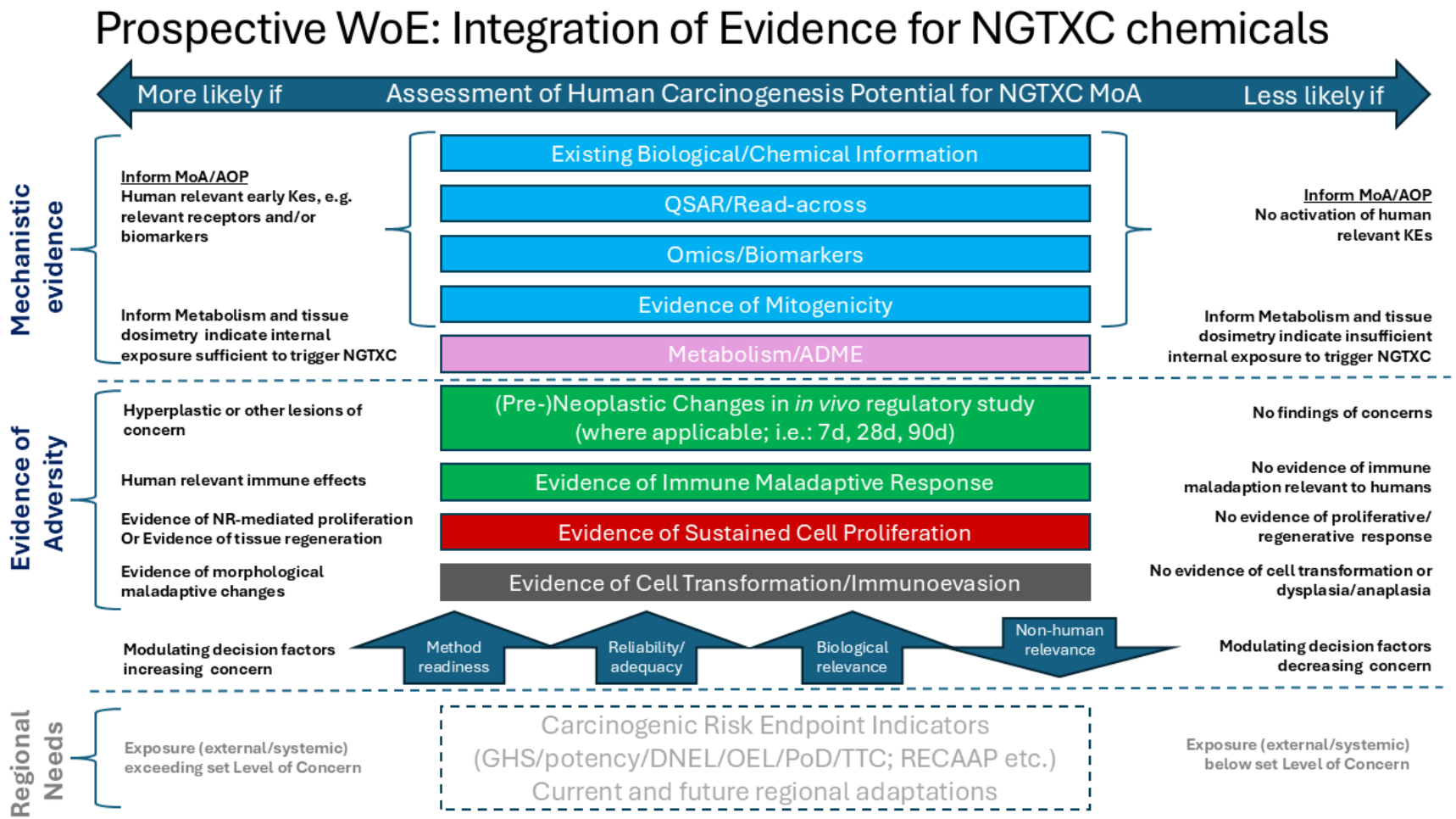


Figure 9. Integration of evidence for non-genotoxic chemicals developed by the OECD NGTXC IATA expert group, with inputs from the OECD NGTXC IATA, ICH S1B (R1) Guideline and ReCAAP framework (ref. OECD ENV/CBC/WRPR(2025)4)*. AOP: adverse outcome pathway; DNEL: derived no-effect level; GHS: globally harmonized system; IATA: Integrated Approach to Testing and Assessment; MoA: mode of action; NGTXC: non-genotoxic carcinogen; NR: nuclear receptor; OEL: occupational exposure limit; PoD: point of departure; TTC: threshold of toxicological concern*

Change Management Requirements for Implementation

Although science-based considerations are the basis for a new approach to carcinogenicity assessment, change management requirements need to be defined in addition and put in place early on for an efficient implementation of such an approach. Elements for this, aside from the **science-based initiatives** thus far mentioned, include addressing the **policy roadmap**, instituting **best practices** and ensuring **communication** to all stakeholders, as depicted in Figure 10. A brainstorming exercise further elucidated the sub-elements shown in Figure 10 to help define relevant actions towards change management in support of the implementation of the new approach for carcinogenicity assessment.

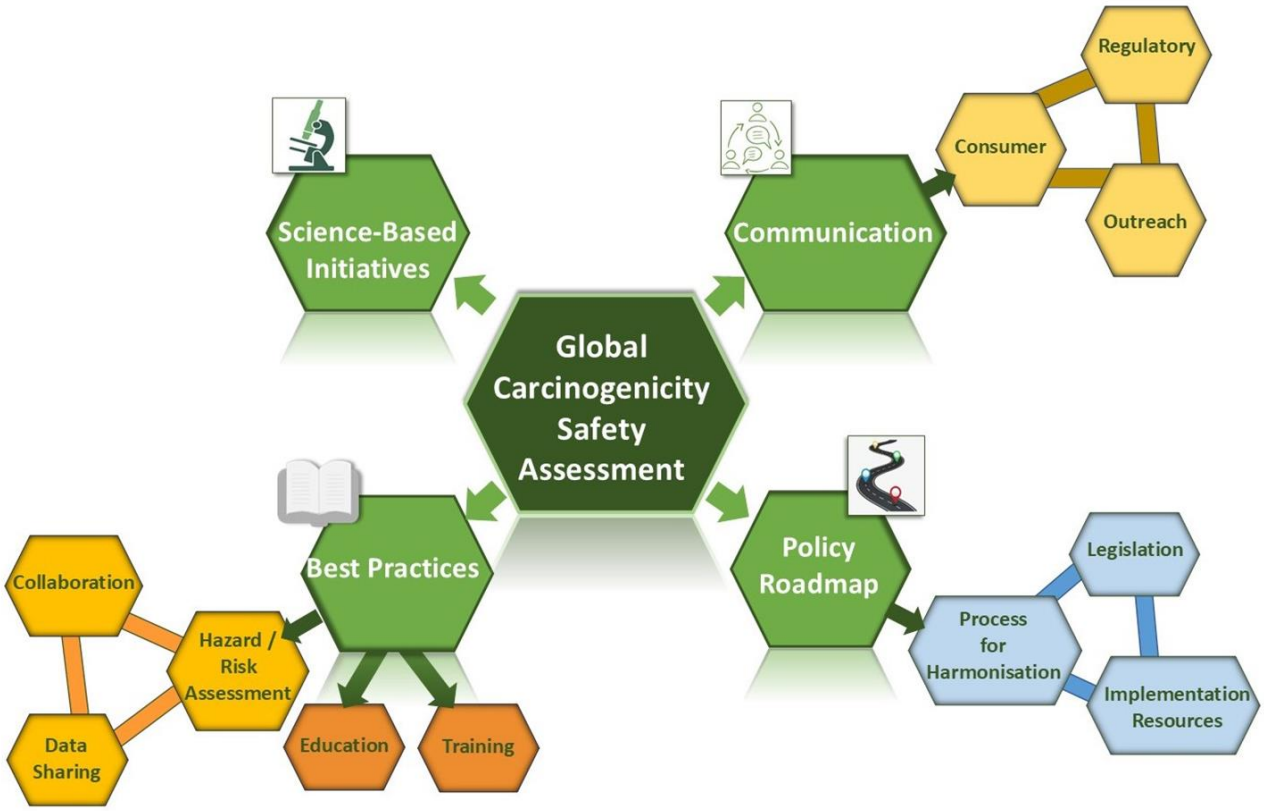


Figure 10. Overview of the different elements that are required for the implementation of a new approach for carcinogenicity assessment.

Summary Action Plan for Implementation of Framework

Presented below are key actions for the short, medium and long-term, presenting requirements for a phased transition towards implementation of the new approach and framework for animal-free carcinogenicity assessment.

Short-Term Actions

- Achieve buy-in from Member States authorities to work towards a new approach for carcinogenicity assessment

- Define realistic protection goals for this endpoint
- Explore adaptation of CLP/GHS criteria using non-animal approaches for carcinogenicity assessment
- Achieve consensus on MOAs relevant for cancer in humans
- Build consensus on the IATA core evidence needed for (non-)carcinogenic substances
- Define and prioritize combinations of non-animal approaches to be used for screening of pivotal, recognized key event perturbation of pathways to cancer
- Establish and test using case studies WoE-based approaches combining non-animal approaches and repeated-dose toxicity studies for different regulations and contexts of use
- Achieve general consensus for a framework for carcinogenicity assessment (see Figure 10) and human relevance across sectors
- Conduct read-across case studies, designed in co-creation with stakeholders, to increase the acceptance of read-across for carcinogenicity assessment
- Improve and refine internal TTC approach and perform case studies to demonstrate in what contexts it could be of added value
- (Further) develop more complex non-animal approaches for late key events of pathways to cancer to enable discrimination between adaptive and adverse responses

Medium-Term Actions

- Perform uncertainty assessment of non-animal approaches for carcinogenicity assessment and compare to traditional approach
- Validate and/or qualify combinations of non-animal approaches to be used for screening of pivotal, recognized key event perturbation of pathways to cancer
- Validate/or qualify more complex non-animal approaches for late key events of pathways to cancer to enable discriminating between adaptive and adverse responses
- Implement new WoE based non-animal approaches for carcinogenicity assessment when deemed acceptable for use in a given regulatory context
- Continue exploring adaptation of CLP/GHS criteria in line with -methods based on non-animal approaches - consider modification of hazard classes for mutagenicity and carcinogenicity
- Review available complex non-animal approaches relevant for carcinogenesis and (further) develop more complex non-animal approaches where needed, to allow for discrimination between adaptive and adverse responses
- Strengthen the gathering of biomonitoring or clinical trial data to enhance exposure knowledge and ensure good quality and clear human data
- Discuss sector-specific case studies to optimize data interpretation and methodologies with relevant stakeholders

Long-Term Actions

- Implement new adapted regulatory process to an animal-free WoE-based carcinogenicity assessment when deemed acceptable for use in a given regulatory context; incorporate internal TTC approach where applicable
- Implement modified CLP/GHS criteria and possibly also modified hazard classes for mutagenicity and carcinogenicity

Conclusion

Jacobs et al. (2016), Luijten et al. (2020) and Hilton et al. (2023) presented objectives, opportunities and challenges with the current regulatory landscape for adopting a WoE-based animal-free carcinogenicity assessment, and raised a call to experts from government, industry, academia and NGOs to convene as an international forum to develop a structured framework for regulatory decision-making that would enable the replacement of the rodent cancer bioassay. With this in mind, and a key part of consensus building, a collaborative approach was undertaken under the umbrella of the EPAA, available initiatives were mapped, a consensus framework was developed, and a plan of action is proposed towards a harmonized strategy for achieving a WoE-based carcinogenicity assessment without long-term rodent bioassays.

The phases and completion of the actions summarized here rely on a continuum of progress from short-to mid- to long-term achievable milestones for policy, science, best practices and communication. The collective goal is to implement an animal-free approach successfully for carcinogenicity assessment at a global level across all sectors. This will be achieved in a stepwise manner. Removing the use of the rodent bioassay will initially necessitate incorporating, where still necessary, short-term *in vivo* (toxicity) tests, while in the long-term the intention is to move towards an animal-free approach, using non-animal approaches with a transparent WoE assessment.

References

European Commission. 2025. Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. Official Journal of the European Union. Available at: http://data.europa.eu/eli/reg/2008/1272/2022-12-17

EFSA Scientific Committee, Bennekou SH, Allende A, Bearth A, Casacuberta J, Castle L, Coja T, Crépet A, Halldorsson T, Hoogenboom LR, Knutsen H, Lambré C, Nielsen S, Turck D, Civera AV, Villa R, Jokelainen P, Zorn H, Benford D, Fitzgerald R, Meek B, Mortensen A, Sand S, Testai E, Manini P, Paini A, Schrenk D. 2025. Statement on the use and interpretation of the margin of exposure approach. EFSA Journal, 23(8), e9606. https://doi.org/10.2903/j.efsa.2025.9606

Heusinkveld H, Braakhuis H, Gommans R, Botham P, Corvaro M, van der Laan JW, Lewis D, Madia F, Manou I, Schorsch F, Wolterink G, Woutersen R, Corvi R, Mehta J, Luijten M. 2020. Towards a mechanism-based approach for the prediction of nongenotoxic carcinogenic potential of agrochemicals. Crit Rev Toxicol. 50(9):725-739. doi: 10.1080/10408444.2020.1841732

Hilton GM, Corvi R, Luijten M, Mehta J, Wolf DC. 2023. Towards achieving a modern science-based paradigm for agrochemical carcinogenicity assessment. Regul Toxicol Pharmacol. 2023 Jan;137:105301. doi: 10.1016/j.yrtph.2022.105301.

Hilton GM, Adcock C, Akerman G, Baldassari J, Battalora M, Casey W, Clippinger AJ, Cope R, Goetz A, Hayes AW, Papineni S, Peffer RC, Ramsingh D, Williamson Riffle B, Sanches da Rocha M, Ryan N, Scollon E, Visconti N, Wolf DC, Yan Z, Lowit A. 2022. Rethinking chronic toxicity and carcinogenicity assessment for agrochemicals project (ReCAAP): A reporting framework to support a weight of evidence safety assessment without long-term rodent bioassays. Regul Toxicol Pharmacol. 2022 Jun; 131:105160. doi: 10.1016/j.yrtph.2022.105160.

ICH. 2023. ICH S1B(R1) Guideline on testing for carcinogenicity of pharmaceuticals, Step 5. EMA/774371/2022.

Jacobs MN, Colacci A, Louekari K, Luijten M, Hakkert BC, Paparella M, Vasseur P. 2016. International regulatory needs for development of an IATA for non-genotoxic carcinogenic chemical substances. ALTEX. 2016;33(4):359-392. doi: 10.14573/altex.1601201.

Jacobs MN, Colacci A, Corvi R, Vaccari M, Aguila MC, Corvaro M, Delrue N, Desaulniers D, Ertych N, Jacobs A, Luijten M, Madia F, Nishikawa A, Ogawa K, Ohmori K, Paparella M, Sharma AK, Vasseur P. 2020. Chemical carcinogen safety testing: OECD expert group international consensus on the development of an integrated approach for the testing and assessment of chemical non-genotoxic carcinogens. Arch Toxicol. 94(8):2899-2923. doi: 10.1007/s00204-020-02784-5.

Luijten M, Corvi R, Mehta J, Corvaro M, Delrue N, Felter S, Haas B, Hewitt NJ, Hilton G, Holmes T, Jacobs MN, Jacobs A, Lamplmair F, Lewis D, Madia F, Manou I, Melching-Kollmuss S, Schorsch F, Schütte K, Sewell F, Strupp C, van der Laan JW, Wolf DC, Wolterink G, Woutersen R, Zvonar Z, Heusinkveld H, Braakhuis H. 2020. A comprehensive view on mechanistic approaches for cancer risk assessment of non-genotoxic agrochemicals. Regul Toxicol Pharmacol. 118:104789. doi: 10.1016/j.yrtph.2020.104789.

Louekari K, Jacobs MN. 2024. A modular strategy for the testing and assessment of non-genotoxic carcinogens. Arch Toxicol. 98(8):2463-2485. doi: 10.1007/s00204-024-03753-y.

OECD. 2018a. Test No. 451: Carcinogenicity Studies, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264071186-en.

OECD. 2018b. Test No. 453: Combined Chronic Toxicity/Carcinogenicity Studies, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264071223-en.

4.4. Repeated dose toxicity

4.4.1 EFSA chapter describing where the number of animals & species tested can be reduced

Background

Chemical risk assessment has been built based mainly on animal testing. These studies aim to predict hazard identification and human safety protection. In several cases, such as in pesticide active substances, regulation sets data requirements to test more than one species, in studies with same duration. In other cases, animal studies are still required where safety information can already be retrieved through different assessment. With evolution of science and scientific knowledge, the strategies used for safety assessment might be revised and where possible updated.

The OECD guidance on the use of research data[70], in the discussion of regulatory relevance highlights that "Regulatory relevance changes as the regulatory framework develops over time. What is not relevant in the current framework may become relevant in the future, as new legislation or scientific guidance is established and *vice versa*." Relevance can change depending on the mandatory data requirements, but the later ones can also be adapted to the studies that are (over time) considered less relevant to ensure human safety protection.

The collective knowledge and data along the years that the animal studies have built (requests dates back to 1960 (Box & Spielmann, 2005)) allow an informed retrospective analysis to evaluate where animal information might be redundant or unnecessary into human protection decisions. In 2022, EFSA started to explore the relevance of dog studies for the derivation of health-based guidance values for plant protection products (PPP) approval (Panzarea et al., 2022). The process involved the building of a dedicated database, retrospective analysis followed by experts' judgment. A similar process can be applied for other animal studies.

Although such process should be done continuously according to the Directive 2010/63/EU on the protection of animals, the European Commission roadmap is the opportunity to revise strategies that are mainly based on animals and might not be the best approach according with the current scientific knowledge and innovation.

Ongoing actions

Waiving of Sub-Chronic (90-day) dog studies with further aim to phase out

For the approval of plant protection products (PPPs), the scientific rationale of using the dog as 'second' species in the regulatory process has been debated for long time and culminated in the elimination of the one-year dog study (OECD TG 452; 2018) from the data requirements in the European Union (EU).

Under Regulation (EU) 283/2013 on setting out the data requirements for pesticides active substance, short-term oral toxicity testing in rodents (90-day rat study) and non-rodents (90-day dog study) species are still required to address hazard identification and human safety of PPPs and to support the active substance approval in the EU.

In 2022, the EFSA PREV Unit and Plant Protection Products and their Residues (PPR) Panel members reviewed the available data on dog studies conducted with the pesticide active substances previously marketed in the EU and/or currently approved by the Union legislation. A retrospective analysis of the results on setting of health-based guidance values (HBGVs) i.e., acceptable daily intake (ADI), was performed.

[70]OECD Guidance Document on the Generation, Reporting and Use of Research Data for Regulatory Assessments Series on Testing and Assessment No 417, 31 Oct 2025, pdf

In 101 out of 400 active substances registered in Europe, dog studies were used to set the ADI, 96% of which experimental and biological variability, allometric body weight scaling and dose spacing were able to explain the differences observed between rodent and non-rodent species. Only for 5 active substances the dog 90-day study showed to be the most relevant for risk assessment without the dismissal of the factors listed previously. There was, however, no cases where dog studies showed to include endpoints of toxicity based on physiological similarities with human. The expert analysis also concluded that for the cases where dog seemed to be the most sensitive species would need further additional evidence.

As a follow up, a new project is ongoing (following a new mandate[71]) that aims to continue contributing to the debate by reviewing/updating existing data to develop a tiered assessment strategy, recommending the possible waiving of dog studies in the regulatory process when deemed scientifically justifiable.

The working group (WG) constituted of EFSA experts is revising the data collected with the previous project to build the tiered approach. Before finalising, the group aims to:

- Test the tiered assessment strategy with proof-of-concept case studies with experimental data provided by industry;
- Consult with relevant stakeholders (through a workshop) to refine and/or further improve the strategy
- Harmonise with other ongoing projects (please see in next section, Recommendations from EPAA project team for phasing out the non-rodent species testing)

Overall, the project can be the basis for a future removal of the 90-day dog study as second species from PPP data requirements. But most of all, can be used as an example of the type of analysis that can be done to reconsider the need of other (90-day) animal studies.

EFSA recommends that the same exercise to be done for animal studies starting from the following questions:

A. What is the value of a by default 'second' species when all are a surrogate for humans?

B. What is unique of the study to be of any benefit in the chemical risk assessment and provide a protective ground for human population?

C. The group of animals (e.g. number/sex/group) treated really covering the intra- and interspecies variability aspects?

D. Taking into consideration all 90-day studies available, how many play a pivotal role for hazard identification and/or setting protection levels? Could the information from these studies used to set protection levels, be retrieved from other studies?

E. What is the impact of the study on the targeted protection level to be achieved for the exposed human population, accepting that the study is specific enough or are there other approaches more suitable?

90-day study for enzymes

Following a similar principle as for the 90-day dog study, EFSA is carrying out a retrospective analysis regarding the use of 90-day rat study for enzyme testing.

Under regulation (EU) No 234-2011 setting out the data requirements for food enzymes, sub-chronic oral toxicity testing in rodents (90-day study) is required for human safety evaluation of food enzymes and to support their approval in the European Union.

[71] https://open.efsa.europa.eu/questions/EFSA-Q-2024-00199

Under the current EFSA Guidance (EFSA CEP PANEL, 2021), food enzymes produced from animals or plants commonly consumed as food, do not require additional *in vivo* toxicological data. Moreover, waiving of 90-day tox studies for food enzymes from microbial sources is already contemplated in the current guideline for the following cases:

i) the criteria for a QPS (Qualified Presumption of Safety) are met for a production microorganism and there is no concern from the manufacturing process,
ii) there is no or negligible carry-over of the FE-Total Organic Solids (TOS) into the final product or
iii) an appropriate substitute of toxicological data is available (EFSA CEP Panel, 2021).

Following the approval of a self-mandate, a new WG of Experts has been established to review data available from food enzyme dossiers submitted to date. The aim of this retrospective review is to assess the hazard identification capacity of repeated 90-day oral studies in the food enzyme domain and to identify those cases where waiving of studies in the regulatory process would have been scientifically justifiable.

Moreover, the use of non-animal approaches that could reliably and efficiently produce information that ensures robust safety assessment of food enzymes to human consumers while reducing animal use will be explored.

Development and use of guidance documents that rely on tiered approaches to minimise the use of animals and maximising the use of Non-Animal Methods

Despite alternatives to animal testing being strategically enforced in almost every legislation/regulatory area, many of them still rely in animal studies. While the complete phase out or replacement of animal studies is still not scientifically (or regulatorily) possible, an approach that prioritises the use of non-animal approaches to retrieve the relevant information should be adopted. A sector-specific tiered approach should be developed that relies on non-animal approaches in the lower tiers, that could distinguish chemicals with potential harm from non-toxic ones, understand its ADME and mechanistic properties of the chemical but also possible species differences. With this initial approach, specific and defined triggers could be incorporated into the lower tier that would help to decide where additional animal studies would be necessary or where animal studies would not be relevant. This approach gives more space to accommodate non-animal approaches and accompanied its development and availability, pushing animal studies to higher tiers and gradually reducing their need.

The implementation of a tiered approach is more challenging in regulatory areas where data requirements are specified in the legal text, which is a minority of the legislative areas under the EFSA remit (PPPs, GMO, Feed additives and Food enzymes). Nonetheless, in the Feed additives area, despite the Regulation (EC) No 429 establishes the requirements for the risk assessment, the EFSA FEEDAP Panel adopted guidance documents which are used in the assessments. With the years, waivers, use of existing data, extrapolations of conclusions between species, and use of in vitro tests have been applied. This has resulted in an effective reduction of the studies needed in our assessments. The same occurs in the Food enzyme domain where data requirements are established in Regulations (EC) No 234/2011 and No 562/2012, and the EFSA CEP Panel Guidance is followed in the assessment process.

For Food Enzymes and Food Additives and Flavourings, data requirements in all three areas are established in Regulation (EU) No 234 and specific Guidance documents have been developed by EFSA. In other sectors such as Novel Foods and Food Contact materials, the process is led by self-mandated guidance's coordinated and approved by EFSA panels that are renewed regularly. This process makes the update of risk assessment procedure easier, and more prone to accommodate the latest scientific technology, especially if relying into a tiered approach.

In September 2024, EFSA has published the updated scientific guidance for the preparation of applications for authorisation of novel foods[72]. In the revision process the Nutrition, Novel Foods and Food Allergens (NDA) updated the Tiered approach to conducting toxicology studies (Figure 11) that reinforce the use on alternatives to animal testing and available data before conducting any animal study. The current approach, put more emphasis on ADME at lower tiers and allows the early identification and screening of species differences that would highlight the unnecessary use of animal studies (ex.: rapid and complete degradation in the GI tract into components with an established safety profile)

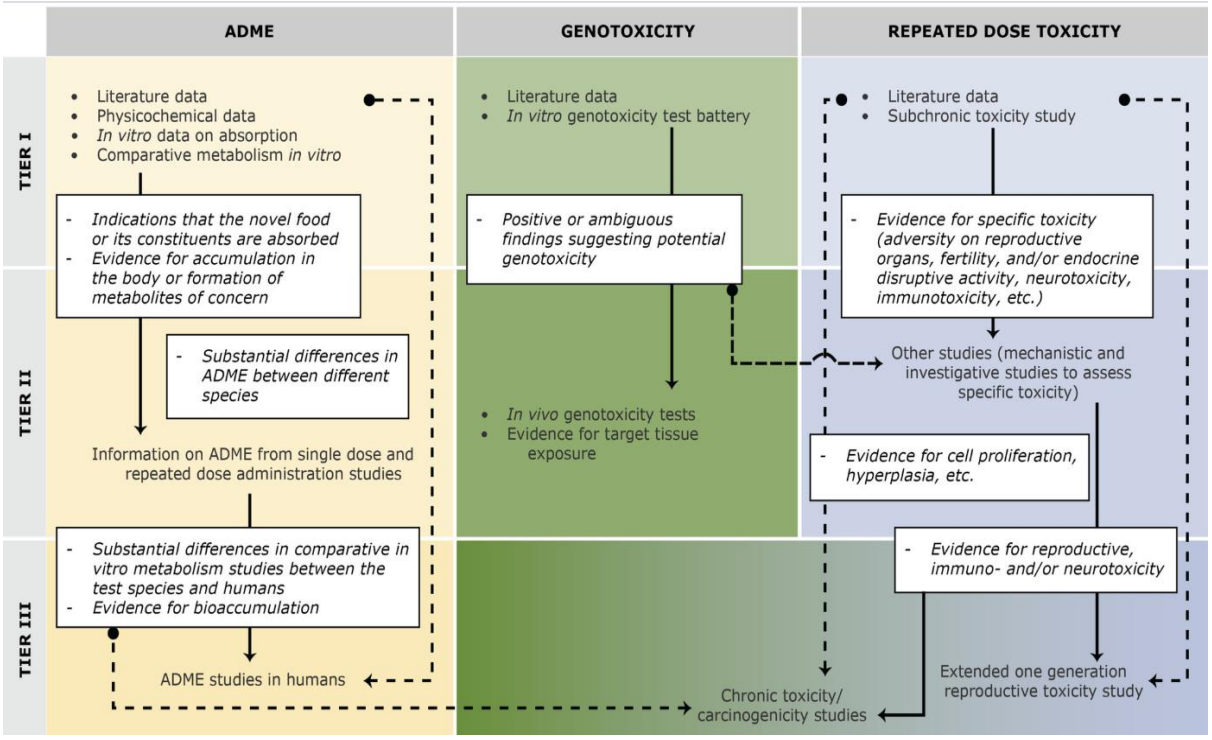


Figure 11. Illustrative example of the tiered approach from EFSA guidance on the scientific requirements for an application for authorisation of a novel food in the context of Regulation (EU) 2015/2283 https://www.efsa.europa.eu/en/efsajournal/pub/8961 | *"Tiered approach for ADME & toxicological assessment of novel foods. The boxes in white represent triggers that will call for higher-tier studies as indicated by the arrows. Solid lines refer to the standardised way of conducting a tiered-based approach by moving from lower to higher tier assessments; dashed lines refer to a deviated tiered-based approach where the existing data require Tier III studies, from ADME to toxicity or from genotoxicity to repeated dose toxicity"*

EFSA's Panel on Food Additives and Flavourings (FAF) has been working on the revision of the "Food Additive Guidance for submission for food additives evaluation" (EFSA ANS Panel, 2012) since October 2023, following a self-task mandate to update the content of the guidance in the light of the experience gained with the safety assessment of food additives applications submitted under Regulation (EC) No 1331/2008.

A first draft of the revised guidance was endorsed by the FAF Panel in November 2024 which was published for public consultation on from December 2024 until February 2025. During the public consultation several comments were received concerning the proposed toxicological data requirements and the revised tiered approach for toxicokinetic and toxicity testing. It was suggested to consider adapting the study design of toxicity studies and possibly combining multiple toxicokinetic and toxicity endpoints in one study to enhance the efficiency and ethical compliance. It was also suggested to apply a more flexible approach on the possible use of Non-Animal Method-

[72] https://www.efsa.europa.eu/en/efsajournal/pub/8961

based methods in the risk assessment, even if not yet validated, to be considered as part of a weight of evidence approach.

The Panel underwent a comprehensive revision of the tiered approach for toxicity studies, in view of the comments received. The updated version of the tiered approach now recommends prioritising the use of existing information and of non-animal testing (in silico and in vitro) in the first Tier to inform on the safety of the proposed food additive. When *in vivo* animal studies are recommended in higher Tiers, in some cases study designs have been modified with the aim of minimizing the use of experimental animals and maximizing the toxicological endpoints under investigation, in line with the 3Rs principle. The revised food additive guidance was adopted by the FAF Panel in November 2025 and was published in January 2026[73].

The assessment of the safety and efficacy of the feed additive, done by the FEEDAP Panel, is done in line with the guidance documents produced by the FEEDAP Panel. The first set of guidance documents was released in 2008 and were in line with the implementing rules of Regulation (EC) No 1831/2003 available in Regulation (EC) No 429/2008. The FEEDAP Panel guidance documents have been updated with the knowledge gained in the assessment of feed additive and the data requirements, especially regarding animal studies, have been reduced in comparison to the ones established in the implementing regulation. At this regard, the newest guidance documents[74] establish

i) the possibility to use available data (e.g., for ADME, toxicological studies, safety/efficacy for the target species),
ii) more waivers for data considering the type of substance,
iii) purity and/or origin, and increased possibilities to extrapolate results between species, and iv) the use of in vitro tests.

Examples in use already include (not exhaustive list):

- Assessment of flavouring substances using already existing data for ADME and/or NOAELs, read-across between substances,
- To establish the safety for the target species the applicants had to conduct tolerance trials in the different target species. Nowadays, one single study in a laboratory animal can be used to conclude on all animal species for most of the substances under evaluation. The applicants may also make use of already existing data[75].
- The number of efficacy trials has been reduced considerably by increasing the number of possibilities to extrapolate data[76],
- In the case of the sub-chronic oral toxicity study (section 3.1 and 3.2.3 in the guidance on the safety for the consumer[77]), the waivers applied in the guidance on the safety for the consumer, have reduced considerably the number of dossiers in which the sub-chronic oral toxicity study (and genotoxicity testing) are needed.[78] Major impact has been seen in fermentation products (e.g. enzymes and amino acids).

The guidance documents are revisited periodically to see further possibilities to reduce-refine-reuse data.

Other possible actions/recommendations to be explored by EFSA

[73] Administrative guidance for the preparation of applications on food improvement agents (food enzymes, food additives and food flavourings)
[74] Available at https://www.efsa.europa.eu/en/applications/feedadditives/regulationsandguidance; safety for the target species, safety for the consumer, safety for the user, safety for the environment and efficacy guidance.
[75] https://www.efsa.europa.eu/en/efsajournal/pub/5021
[76] https://efsa.onlinelibrary.wiley.com/doi/10.2903/j.efsa.2024.8856
[77] https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2017.5022
[78] https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2017.5022

Refinement action

Following the tiered approach, the requirement for animal studies is generally intrinsic to higher tiers and therefore, their removal is often seen as impossible. In such cases, it is important the maximization of endpoints while maintaining or minimizing the number of animals. When *in vivo* animal studies are recommended in higher Tiers, study designs can be modified in some cases to allow for more information to be retrieved from the *in vivo* study. The aim is minimizing the use of experimental animals and at maximizing the toxicological endpoints under investigation, in line with the 3Rs principle.

In the guidance on the assessment of the safety of feed additives for the target species[79] the EFSA Panel on Additives and Products or Substances used in Animal Feed introduced the possibility of extrapolating the results (NOAEL/BMD) derived from toxicological studies with laboratory animals to the different species/categories of animals based on simple calculations (conversion of mg/kg bw and day to mg/kg feed). This approach implied a considerable reduction in the number of tolerance studies required, especially for those substances in which 90-day toxicity studies were needed.

Similar considerations and approaches can be adopted in other type of assessments.

90-day rat and 28-day animal studies for Genetically Modified Organisms (GMO) assessment

Under the current regulatory framework for GMOs (Regulation (EU) No 503/2013), applicants are mandated to submit a 90-day feeding study in rats using the whole food/feed derived from GM plants.
However, the EFSA GMO Panel considers that such animal studies are not routinely necessary, as noted in its opinions on GM food and feed. Instead, they should only be considered on a case-by-case basis. According to EFSA's experience, the decision to conduct an animal study should be based on the results of molecular characterisation, comparative analysis, and toxicological assessment. If these evaluations do not reveal concerns, animal studies should not be required.[8081]
Given the GMO Panel's position on 90-day studies and EFSA's ongoing work in related areas, this topic could benefit from expert analysis and discussion. It could contribute to a reflection on whether to waive the legal requirement for 90-day oral toxicity studies in rats for whole GM food/feed assessment.

Furthermore, the EFSA GMO Panel recently published a scientific opinion addressing current practice, challenges and future opportunities in the safety assessment of proteins in genetically modified organisms (GMOs) (EFSA GMO Panel, 2025)[82]. The Panel conducts safety assessments of proteins in GMOs based on its established guidance documents and relevant Codex Alimentarius guidelines, including the use of animal studies. These approaches have effectively supported evaluations over the past two decades. However, the field is evolving. There is an increasing need to reduce reliance on animal testing, and the growing complexity of GMO applications calls for a modernised assessment strategy. Traditional toxicological *in vivo* studies were originally developed for simple chemicals, and their relevance for assessing proteins remains uncertain. In its latest opinion, the GMO Panel recommends a stepwise, risk-based approach in which *in vivo* studies are conducted only when potential hazards are identified or when essential safety information cannot be obtained through alternative methods (see Figure 3 in EFSA GMO Panel, 2025). In short, an

[79] https://www.efsa.europa.eu/en/efsajournal/pub/5021
[80] EFSA Panel on Genetically Modified Organisms (GMO); Scientific Opinion on Guidance for risk assessment of food and feed from genetically modified plants. EFSA Journal2011;9(5): 2150. [37 pp.] doi:10.2903/j.efsa.2011.2150.
[81] EFSA (European Food Safety Authority), 2014. Explanatory statement for the applicability of the Guidance of the EFSA Scientific Committee on conducting repeated-dose 90-day oral toxicity study in rodents on whole food/feed for GMO risk assessment. EFSA Journal2014;12(10):3871, 25 pp., doi:10.2903/j.efsa.2014.3871
[82] https://efsa.onlinelibrary.wiley.com/doi/full/10.2903/j.efsa.2025.9568

improved strategy for protein safety assessment underlined in this document includes: (1) considering history of safe use (HoSU), read-across and phylogeny defining the type of data required and remove the need for specific in vitro or *in vivo* studies; (2) applying advanced in silico tools, including predictive computational models and improved phylogenetic analysis to enable more accurate comparisons with known allergens, toxins or 'safe' proteins; (3) using standardised in vitro gastrointestinal models that replicate physiological conditions; (4) developing targeted *in vivo* studies following a hypothesis driven approach; (5) evaluating the role of exposure in the safety assessment; and, where necessary, (6) considering post-market monitoring for risk characterisation. The EFSA GMO Panel aims to integrate alternatives to animal testing into risk assessment to provide a scientific basis for waiving *in vivo* testing, aligning its approach with the 3Rs principles (Replacement, Reduction, Refinement) and the European Commission's roadmap for phasing out animal testing.

References

Box, R.J., Spielmann, H. Use of the dog as non-rodent test species in the safety testing schedule associated with the registration of crop and plant protection products (pesticides): present status. Arch Toxicol 79, 615–626 (2005). https://doi.org/10.1007/s00204-005-0678-0

Panzarea, M., Terron, A., Coja, T., Pelkonen, O. Editorial: Relevance of dog studies for the derivation of health-based guidance values for plant protection products approval (2022). https://doi.org/10.2903/j.efsa.2022.e200923

EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP); Guido Rychen, Gabriele Aquilina, Giovanna Azimonti, Vasileios Bampidis, Maria de Lourdes Bastos, Georges Bories, Andrew Chesson, Pier Sandro Cocconcelli, Gerhard Flachowsky, Jürgen Gropp, Boris Kolar, Maryline Kouba, Marta López-Alonso, Secundino López Puente, Alberto Mantovani, Baltasar Mayo, Fernando Ramos, Maria Saarela, Roberto Edoardo Villa, Robert John Wallace, Pieter Wester, Montserrat Anguita, Jaume Galobart, Matteo Lorenzo Innocenti, Laura Martino. Guidance on the assessment of the safety of feed additives for the target species (2017). https://www.efsa.europa.eu/en/efsajournal/pub/5021

4.4.2 Chapter from EPAA project team with recommendations for phasing out the non-rodent species testing for (sub-) chronic toxicity testing with alternatives to animal testing across sectors

Recommendations

Basket 1: Short term

- **Legal framework and regulatory use mapping:**
  - Conduct a comprehensive landscape mapping of the different regulatory requirements and uses of non-rodents in repeated-dose toxicity testing across the different sectors.
  - Identify where it would be possible to leverage 3Rs approaches that are already implemented to promote cross-sectorial alignment.
- **Tiered approach development:**
  - Integrate international work in non-animal approaches and EFSA and US EPA evaluations of the dog study to develop a waiver framework and tiered strategy aimed at the elimination of dog testing for agrochemicals.
  - Conduct an analysis to translate the process and outcome to other industry sectors like feed additives, biocides and veterinary medicines.
  - Introduce tiered approach to evaluate the need for two species (e.g. rodent and non-rodent) chronic toxicity testing for human medicine.
- **Animal reduction and refinement:**
  - Consider study design optimisation to reduce and refine the use of animals in the chronic and/or precursor dose-range finding studies in all sectors.
  - EU-level and national legal framework mapping on refinement during in-life in the laboratory of large animals
  - Conduct a review of EU-level and national policies for rehoming and map where there would be opportunities for rehoming (e.g. non-terminal procedures).
- **Integration of advanced new approach methodologies comparative systems**:
  - Explore the use of comparative rodent and non-rodent and/or human ADME/ toxicokinetic in silico and in vitro systems and toxicodynamic systems to enhance the understanding of species sensitivity, species translational capacity and human relevance which can guide the selection of appropriate representative species.
- **Global collaboration:**
  - Strengthen existing partnerships and promote global stakeholder to achieve alignment on reducing the use of non-rodent as a second species.
  - Facilitate international harmonisation through activities like the International Medicines Regulators' Working Group on 3Rs. Create a multistakeholder 3Rs group, with representatives from each sector. Allow these to engage in agenda setting, research priorities where all stakeholders/sectors benefit (e.g. Regulatory Science Network Netherlands or HESI).

Basket 2: Medium term

- **Standardisation and validation of in vitro methods**:
  - Promote standardisation and validation according OECD GD34 of in vitro assays to assess comparative ADME/toxicokinetic and toxicodynamic in rodent and non-rodent species and/or human systems.

- **Regulatory endorsement and industry incentives:**
  - Promote regulatory acceptance and provide industry incentives via relevant European agency qualification advice on Non-Animal Method protocols and methods development for humand and veterinary medicinal products.
- **Waiver framework expansion:**
  - Promote industry incentives for adoption of tiered approaches to industry to eliminate dog use in this sector.
  - Promote broader international adoption of waiver framework in the agrochemical sector and tiered approach in other regulated products with regards to non-rodent testing.
- **Innovative research and data sharing:**
  - Foster the research and development of new approaches such as virtual control group (e.g. VICT3R), to further reduce non-rodent use.
  - Use and reporting of historical control data (HCD) methodology (EFSA, 2025[83]) which supports refinement and reduction through strengthening the statistical power of existing toxicity studies and avoiding additional confirmatory studies. This complements the above virtual control groups initiative, as the HCD methodology is relevant for assessing the adequacy and quality of control datasets under development.
  - Promote cross-sector data sharing initiatives and link them to funding mechanisms to accelerate progress.
- **Legislative adaptation**:
  - Explore and propose updates to EU legislation to accommodate tiered and alternative, non-animal testing approaches.

Basket 3: Long term

- **Advance research on non-animal alternatives**:
  - Further research and development of the non-animal approaches such as computational methods explored in the NC3Rs CRACK IT Challenge 'Virtual Second Species', to replace animal studies.
- **Global regulatory adaptations:** Substantial progress will require international legislation and guideline adaptations to enable tiered and alternative testing approaches and ultimately ensure removal of the non-rodent species.

Background

Testing a second, non-rodent species is a requirement in several industrial sectors in the EU and beyond. This testing is commonly conducted in dogs for (sub-) chronic toxicity and rabbits for developmental toxicity in addition to rodent studies. Nevertheless, other species such as mini-pigs and non-human primates could also be used for human medicines. The rationale for conducting non-rodent studies is to identify any potential health hazards that would not have been detected in rodent studies alone. However, alongside issues with translation of data from animals to humans, the experimental design of non-rodent studies has several limitations, including low statistical power of the studies, leading to uncertainties in interpretation. As a consequence, the added value of these studies as part of a weight-of-evidence approach is questioned. In addition, recent analyses have shown that, for agrochemicals at least, testing in the dog as the second species, provided little added benefit to most risk assessments. (Bishop et al., 2023, Prior et al., 2020)

[83] Use and reporting of historical control data for regulatory studies - - 2025 - EFSA Journal - Wiley Online Library

In its current state, the recommendations focus on the non-rodent species testing for sub-chronic toxicity and/or chronic toxicity studies which are typically required in several sectors (e.g. agrochemical, biocidal, feed additive, human and veterinary medicinal products). The different industry sectors are starting from different states of requirements and readiness for phasing out dog studies. An overview of the information requirements for each sector is presented in Table 8 hereafter. In certain sectors, such as agrochemicals, comprehensive analyses of registered active substances have demonstrated that for the vast majority of pesticides the required 90-day dog study does not provide additional insights relevant to human health protection beyond the required toxicological data package (e.g. rodent studies) submitted as part of the regulatory package. Therefore, efforts should focus on identifying the limited number of substances for which dogs may exhibit greater sensitivity relevant to human. This can be achieved through integration of non-animal approaches through comparative toxicokinetic and toxicodynamic assessments to evaluate any potential implications for human safety. In other sectors, where the relevance of dog studies to human health may differ (e.g. first in human clinical trial starting dose), the development and application of advance predictive tools and non-animal approaches is necessary. These should be tailored to the specific regulatory and scientific context of use of each sector.

In order to phase out non-rodent species, extension of the scope to other sectors such as the food-industry (e.g. food additives) and to other non-rodent second species uses (e.g. developmental toxicity) is needed in the future.

Current information requirements

Table 8: Summary of the regulatory requirements with respect to second non-rodent species for (sub-)chronic toxicity in the different sectors

Industrial sectors	Non-rodent required as a second species for repeated dose toxicity testing Test guideline (TG)?	Is dog always the non-rodent species?	Human health protection goal	Regulatory reference requirements for non-rodent species for repeat dose toxicity	EU Actors
Veterinary medicinalproducts	Non-food producing animals: no except if the dog is the target animal species (a 90-day dog toxicity study can be conducted instead of rodent) Food producing animals: Yes 90-day (OECD TG 409): required Chronic (OECD TG 452): conditional OECD GD 116 as general guidance	Dogs as the most commonly used species (e.g. Swine, mini pigs may also be used)	Dog as the target animal species: - Target-animal safety - Secondary protection objective for human health Food producing animal: - User and consumer/human food hazard and risk assessment	Reg. (EC) No 2019/6 &Delegated Reg. (EC) No 2021/805; .Reg. (EC) No 470/2009. Reg. (EC) No 2018/782 VICH guidance 31, 37, 41, 43, 44	EMA, EC
Human medicinal products	Yes, short-term and chronic (OECD TG 409, TG 452; OECD GD 116) Specific adjustments on study duration and one or two species approach depends on modalities and indications (e.g. biologicals, radiotherapeutics)	No, the most common species is the dog, but the most appropriate species should be used based on pharmacology, target homology, etc. (e.g. dog, NHP, mini pigs)	- Identification of potential target organs (on- and off-target toxicity | exaggerated pharmacology | chemically-mediated toxicity) - Assessment of reversibility of findings observed - Recommendations for clinical monitoring of findings - Exposure margins to intended human (therapeutic) exposure (MoE = margins of exposure) - Recommendation of starting dose for FiH (first-in-human) clinical trial - Product Labelling	Dir. 2001/83/EC & EU reg. 726/2004 ICH guideline M3(R2), S4, S6(R1), S9 Guideline on repeated dose chronic toxicity (CPMP) CPMP/SWP/1042/99 Rev 1 Corr, and duration CPMP/ICH/300/9	EMA, EC
Biocide products	Yes, conditionally requirement (OECD TG 409 if required)	Dogs as the commonly used species	- Hazard identification (C&L) - Consumer (dietary) and non-dietary (industrial, professional, non-professional and secondary [e.g., indirect]) risk assessment	Reg. (EC) No 528/2012	ECHA, EC

Plant protection products	Yes, for conventional active substances. (OECD TG 409; OECD GD 116) Exemption for Microbials.	The dog is the requested species in the regulation	- Hazard identification (C&L) - Consumer (dietary) and non-dietary (worker, operator, bystander, resident) safety (reference dose setting: ADI, AOEL)	Reg. (EC) No 283/2013 specify the requirements for plant protection product active ingredients	EFSA, ECHA, EC
Feed additives	Yes, if necessary 90-day (OECD 409) 1-year (OECD 452)	Dogs as the most commonly used species (Swine, mini pigs may also be used) - OECD 409.	- Target species safety - Consumer safety (ADI, UL) - Users/workers safety	Reg. (EC) No 429/2008	EFSA
Food additives	No. One rodent species (e.g. rat) only. Non-rodent only if the rodent is not a relevant species	N/A	- Consumer safety	Guidance for submission for food additive evaluation (EFSA ANS Panel, 2012)	EFSA, EC
Food contact material	No	N/A	N/A	N/A	EFSA, EC
Industrial chemicals	No	N/A	N/A	N/A	ECHA, EC
Cosmetics	No	N/A	N/A	N/A	SCCS, EC

TG: test guideline, C&L: classification and labelling; Reg.: regulation; UL: tolerable upper intake; ADI: acceptable daily intake

Activities related to the recommended actions

Table 9: Activities identified related to the short:medium:long-term recommended actions

Actions	1st Basket – Improve the existing knowledge base - Short term activities (< ~5 yr)	2nd basket - Replacement after further method development activities	3rd basket - Transition to animal-free regulatory system activities
Mapping Scientific Gaps	(All sectors) Comprehensive landscape mapping of the different regulatory uses of non-rodents in repeated dose toxicity testing across the different sectors and identification where it would be possible to leverage 3Rs approaches that are already implemented in other sectors. *EPAA as a potential actor*		
	(AgChem) Evaluate results from independently conducted retrospective analyses from different stakeholders, in EU and beyond, on the additional value of the 90-day dog test for human health risk assessment for agrochemicals Ongoing projects: regulatory agencies (EFSA [6,9], US EPA), industry (e.g. Syngenta, Corteva)), review articles [1]		
	(Agchem/biocide) Retrospective analysis of dog use for hazard classification and labelling No ongoing project identified. *ECHA or Industry as a potential actor*		
	(Human medicine) Retrospective analysis of the use of two species in (sub-)chronic studies for human medicine Ongoing project: NC3Rs two species project phase 2 [10], review articles [2, 3, 4, 7] *Other potential actor: EMA*		
	(Biocides, veterinary medicines) Retrospective analysis of the value of the dog study in risk assessment No ongoing project identified *Potential actors: EMA, EFSA, ECHA, industry*		

Development and integration of innovative NAMs	(All sectors) Landscape mapping of comparative toxicokinetic in vitro systems in rodent and non-rodent species (e.g. in vitro comparative metabolism study) and target organ systems to investigate comparative toxicodynamics and their application to support species sensitivity, human relevance and species translational capacity. *Potential actors: industry, NAMs developers, CRO*		
	(All sectors) Accelerate validation and standardisation of comparative in vitro ADME methods that could inform PBPK modelling and OECD test guideline development. Example of ongoing projects: OECD test method development proposal for plasma protein binding studies; intrinsic clearance studies *Potential actor: OECD members, ICAPO*		
	(Human medicine) Introduce tiered approach to evaluate the need for two species (e.g. rodent and non-rodent) chronic toxicity testing if conditions allow for human medicine *Potential actors: animal protection NGOs, EMA*		(Human medicine) Application of advances in research into computational and mathematical modelling approaches (e.g. development of a suite of virtual dog tissues to model toxicological endpoints for New Chemical Entities NCE) Ongoing project: NC3Rs CRACK IT Challenge: virtual second species [11] *Actors: NC3Rs, Industry data sharing*
Waiver framework and tiered approach strategy development	(Agchem) Waiver framework development Ongoing project: EFSA [9] (Q4 2025/Q1 2026 Scientific opinion) *Actors: EFSA, Industry*	(Agchem) Publish results of framework development; conduct outreach and promote international acceptance and use of waiver framework for 90-day dog study to ensure global harmonisation and avoid unnecessary testing should some markets still require the study.	

		Actors: trade associations (e.g Asia-Pacific, Brazil); Authorities (e.g. ANVISA, Japan food safety committee)
	(Agchem) HESI joint TEA (Transforming the Evaluation of Agrochemicals) and PBPK waiver framework development for the 90-day Dog study for agrochemicals aiming to define the conditions under which this study may be necessary to inform human health risk assessment [12] *Actor: HESI*	
		(Human and veterinary medicine/biocide) Further research to demonstrate the applicability of extending the Agchem framework to other product sectors where there is a similar context of use, avoiding unnecessary testing. Development of similar regulatory framework applicable to other product sectors/context of use. *No ongoing project* *Potential actors: ECHA, EMA, EFSA*
Consider study design optimisation and best-practices to reduce and refine the use of animals.	(Human medicine) Toxicity study design optimisation in the (sub-)chronic studies required for safety assessment Dose-range-finding optimisation - Optimise study design, for example using microsampling to remove the need for toxicokinetic satellite animals (NC3Rs ongoing projects [13]; - Encourage science-based case-by-case consideration of the use of recovery animals (NC3Rs ongoing project; Sewell et al., 2014; [8, 14]) - Minimise NHP use in drug development (NC3Rs ongoing project; [5]). *Actors: NC3Rs, industry*	(Human medicine) Developing and implementing virtual control groups to further reduce animal use in the *in vivo* toxicity studies that cannot be waived. Ongoing project: VICT3R on virtual control group in the human medicine sector [7] *Actors:*

		(Agrochemicals and veterinary medicine) Remove concurrent control groups and/or reduce the number of concurrent control animals with inclusion of some virtual control animals after further research and data sharing to develop virtual control group in these sectors Potential actors: *NC3Rs, industry trade association*
	EU-level and international legal framework mapping of refinement opportunities during in-life in the laboratory of large animals and rehoming policies following non-terminal procedures. No ongoing projects identified *Potential actors: GRAAL [16], Beagle dogs of Burgundy[17]: White rabbit (for small animals) [18], Individuals (for farm animals) [19], animal protection NGOs, shelter/rescue partners, university*	
Legislative changes[84] and/or guideline updates in the different context of use	(Human medicines) Example of ongoing activities (consultation): -Reflection paper on the current regulatory testing requirements for medicinal products for human use and opportunities for implementation of the 3Rs[20] -Concept paper on revision of the guideline on the principles of regulatory acceptance of 3Rs (replacement, reduction, refinement) testing approaches (EMA/CHMP/CVMP/JEG-3Rs/450091/2012)[21] *Actor: EMA*	
	(Veterinary medicines) Example of ongoing activities:	

[84] The pharmaceutical legislation already allows flexibility and legislative changes are not foreseen or required there.

	-Overview of the current regulatory testing requirements for veterinary medicinal products and opportunities for implementation of the 3Rs - Scientific guideline [22] *Actor: EMA*		
	(Pesticides) Development of a Roadmap for Action on NAMs in Risk Assessment *Actor: EFSA [23]*		
		(All sectors) Provide incentives to industry to encourage the waiver of non-rodent studies, especially when waiver framework is available for agrochemicals. Promote relevant European agency qualification advice team for human medicine on NAMs protocols and methods and relevant advisory European agency team on tiered approach *Potential Actor: EFSA; EMA CHMP qualification [26]*	
		Legislative adaptations for tiered and alternative approaches in EU. *Actor: European Commission, member state forum discussion*	
			Removal of the OECD test guidelines 452, 409 after full replacement of the dog studies *Actors: OECD members*
Reinforce existing partnership and promote collaboration between	(Agrochem) HESI joint TEA and PBPK Committees (US EPA, industry, EFSA, animal protection NGOs) on the development of a waiver framework for the 90-day dog study for agrochemical risk assessment [25] *Actor: HESI*		

international actors	(Human and veterinary medicine) Facilitation of international harmonisation through the activities of the International Medicines Regulators Working Group on 3Rs (IMRWG3Rs) see Terms of Reference (ToR) for the International Medicines Regulators' Working Group on 3Rs [24] *Potential Actors: International medicines regulators WG on 3Rs*

Legend:
White box: activity not initiated
Yellow box: ongoing activity

References

Published papers

1. Bishop et al. 2023. Is the dog study necessary for pesticide toxicity testing? Crit Rev Toxicol 53(4): 207-228.
2. Bower et al.2022. The Dog as a Second Species for Toxicology Testing Provides Value to Drug Development. Int J Toxicol. 2022 Dec;41(6):431-441
3. Chien. et al. 2023. Re-evaluating the need for chronic toxicity studies with therapeutic monoclonal antibodies, using a weight of evidence approach. *Regulatory Toxicology and Pharmacology*, *138*, 105329.
4. Galijatovic-Idrizbegovic et al. 2016. Role of chronic toxicology studies in revealing new toxicities. Reg Toxicol Pharm 82: 94-98
5. Guffroy et al. 2025. Scientific and Regulatory Policy Committee Points to Consider: Proposal and Recommendations to Reduce Euthanasia of Control Nonhuman Primates in Nonclinical Toxicity Studies. *Toxicologic Pathology*, 53(3):01926233241309905.
6. Panzarea et al. 2022. Editorial: Relevance of dog studies for the derivation of health-based guidance values for plant protection products approval. EFSA Journal 2022;20(9):e200923
7. Prior et al. 2020. Opportunities for use of one species for longer-term toxicology testing during drug development: A cross-industry evaluation. Reg Toxicol Pharm 113: 104624.
8. Sewell et al., 2014. Recommendations from a global cross-company data sharing initiative on the incorporation of recovery phase animals in safety assessment studies to support first-in-human clinical trials. *Regulatory toxicology and pharmacology*, *70*(1)): 413-429.

Hyperlinks

9. https://open.efsa.europa.eu/questions/EFSA-Q-2024-00199?search=2024-00199
10. https://www.nc3rs.org.uk/our-portfolio/review-use-two-species-regulatory-toxicology-studies
11. https://nc3rs.org.uk/crackit/virtual-second-species
12. www.Hesiglobal.org
13. https://nc3rs.org.uk/3rs-resource-library/study-designs-pharmaceutical-and-chemical-development;
14. https://nc3rs.org.uk/our-portfolio/reducing-use-recovery-animals
15. https://www.vict3r.eu/
16. https://www.graal-defenseanimale.org/english/
17. https://beaglesofburgundy.org/
18. https://www.white-rabbit.org/rehabilitation/
19. https://www.facebook.com/Individus63
20. https://www.ema.europa.eu/en/overview-current-regulatory-testing-requirements-medicinal-products-human-use-opportunities-implementation-3rs-scientific-guideline
21. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-reflection-paper-current-regulatory-testing-requirements-medicinal-products-human-use-opportunities-implementation-3rs-revision-1_en.pdf
22. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-reflection-paper-current-regulatory-testing-requirements-veterinary-medicinal-products-opportunities-implementation-3rs-revision-1_en.pdf
23. Development of a Roadmap for Action on New Approach Methodologies in Risk Assessment | EFSA
24. Terms of Reference (ToR) for the International Medicines Regulators' Working Group on 3Rs
25. Transforming the Evaluation of Agrochemicals (TEA) - HESI - Health and Environmental Sciences Institut
26. Regulatory acceptance of new approach methodologies (NAMs) to reduce animal use testing | European Medicines Agency (EMA)

Reducing and refining long-term toxicity testing

ECETOC Thought starter: Reduction of animal numbers under the current EU regulatory schemes

Note: ECETOC is a member of the Human Health working group developing actions for this roadmap. The recommendations described here have been endorsed by the Human Health working group.

Key recommendations:

- Implement tiered approach into EU REACH Annex XI (*in silico*, *in vitro*, Smart *in vivo*, regulatory *in vivo* studies)
- Amend REACH Annexes to allow replacing 90-day repeated dose toxicity information requirement by enhanced 28-day studies
- Develop guidance documents at OECD-level for the design of Smart *in vivo* studies
- Conduct further research to reduce study duration and/or animal numbers and to use add-ons in reproductive/developmental toxicity screening studies according to OECD TG422.
- Adopt improved use of kinetic information at all levels (*in vitro* kinetics, *in silico* modeling/PBK/qIVIVE, *in vivo* (Smart) studies with integrated kinetic), to allow a more human relevant assessment and less reliance on animal studies.

The **action steps are laid out for inclusion in the EU COM roadmap to phase out animal testing enabling a reduction through refinement in current animal testing under the REACH regulation.** Reduction of animal testing can already be implemented under the current regulatory paradigm. It is important to recognize that the transition to animal-free approaches may require longer lead times in certain cases, such as difficult-to-test substances. The work supporting the proposal below is currently ongoing as part of the ECETOC Human Health Transformational Program (HH TP), the Staged Assessment and Smart studies Task Forces.

Summary of the approach

ECETOC developed a framework for the use of alternatives to animal approaches in assessing the safety of chemicals in Ball et al. (2022). The ECETOC framework is a tiered approach using the Threshold of Toxicological Concern TTC (Tier 0), *in silico* assessment (Tier 1), *in vitro* assessment (Tier 2) and targeted Smart *in vivo* studies (Tier 3) which includes hazard characterisation, exposure assessment and risk assessment. The framework can be applied to develop and test a hazard hypothesis and to support read-across and grouping. If a decision for classification or safety assessment can be made with confidence at one of the tiers, further testing might be omitted. Therefore, the ECETOC framework can support reducing and refining animal testing during the transition to phase out animal testing.

The framework became one of the drivers behind the idea of the European Partnership for Alternative Approaches to Animal Testing (EPAA) to promote a Designathon which called for ideas and concepts on how non-animal approaches could be used for the potentially new classification system of chemicals for “human systemic toxicity” (EPAA 2023).

As part of the EPAA’s ‘NAM Designathon 2023’ challenge for human systemic toxicity (Worth et al. 2025), ECETOC has further developed an exclusively non-animal-based classification system for “human systemic toxicity” (Doe et al. 2025). While both EPAA and ECETOC share the common goal of advancing alternative methods, ECETOC’s proposal adopts a different approach. Specifically, it focuses on repeated dose toxicity, similar to the assessment carried out in the classification for ‘Specific Target Organ Toxicity – Repeated Exposure’ (STOT- RE) (Ball et al. 2022; Doe et al. 2025).

During the transition period it would be prudent to retain the four major categories of the classification system: general toxicity, carcinogenicity, reproductive toxicity and mutagenicity, within any revisions.

The first area to be revised would be general toxicity, broadly corresponding to STOT-SE and STOT-RE.

Furthermore, a follow-up publication to Doe et al. (2025) is currently being drafted, where the last tier of the original framework for targeted Smart *in vivo* studies (Tier 3) is addressed.

Extended or "Smart" studies can be defined as *in vivo* studies that employ additional or different endpoints from conventional studies. We envision that these *in vivo* studies will be of short duration (e.g. 28 days). Smart 28-day studies will comply with current OECD test guidelines and be enhanced with modern technologies, such as 'omics', to obtain information on both the toxicodynamics and the kinetics of the substance studied (Henriquez et al. 2024).

In addition to the above-mentioned paper focused on Smart *in vivo* studies, another paper is planned to address what the full framework would entail. How each of the parts would work combined and how the use of Smart *in vivo* studies can increase confidence in the characterisation of a hypothesized hazard for the whole assessment.

Method readiness, applicability domain and validation status

Smart studies[85]

One aspect of the tiered testing strategy proposed by ECETOC is to conduct an extended or smart 28-day *in vivo* study to close relevant data gaps regarding STOT-RE classification if the 90-day study is omitted. Therefore, an analysis of a database with substances tested in 28-day and 90-day repeated dose toxicity studies was conducted to identify potential information gaps by comparing the outcomes of 28-day studies to those of 90-day studies. We evaluated the extent to which short-term studies with 28-day exposure can or cannot forecast the target organs and effects relevant for the classification of substances according to STOT-RE.

Our investigations comparing 28- and 90-day studies have shown that information of relevance to STOT-RE classification can be determined by either duration of study. Specific gaps (pathology effects only observed after 90 days duration) could be effectively covered by the additional measurements (omics, specific markers of toxicity) envisaged in Smart *in vivo* studies. The comparison further indicated that enhanced blood analysis for hormones and haematology is helpful to improve the prediction of organ toxicity in 28-day studies for 90-day studies. Blood can also be used for genotoxicity assessment. The Smart study should also provide data for the transcriptional changes in some specific tissues as well as toxicokinetic assessment, to account for potential accumulation of parent compounds or metabolites. To achieve this, blood microsampling technologies will have to be introduced (Chapman et al. 2014). The study design and hence the additional investigations and technologies to be used serve two purposes:

(1) Enhance the sensitivity of the 28-day study to reduce the necessity for a 90-day study (default – to always include) and
(2) based on the results of the *in silico* (Tier 1, section 2.2) and *in vitro* (Tier 2, section 2.3) assessments to allow for a more in-depth investigation of the modes of action identified (bespoke study design), to allow for example the assessment of human relevance and sensitivities.

Therefore, in addition to the general improvements in the study design, it is also envisioned that compound specific parameters and/or intermediate time points could be included, based on the results of Tier-1 and -2 information. An overview of sample types and potential add-on techniques is provided in Table 10. Not all listed technologies can be added to a 28-day study in all combinations, there are

[85] ECETOC's transformational program A framework to incorporate NAMs in regulatory toxicology. Bennard van Ravenzwaay, John Doe, and Lucy Wilmot. O-28, Pg 72, Abstracts of the 52nd EEMGS meeting – ECETOC Session Arh Hig Rada Toksikol 2024;75(Suppl. 1):22-209. https://eemgs.eu/wp-content/uploads/2024/10/Arhiv_75_Supp1-split.pdf

limitations with regards to sampling time points, number of samples, animal welfare as well as compatibility of methodological requirements.

Overall, this would lead to a similar level of knowledge and confidence from the same or fewer animals studied, even more so for potential future waiving of high animal usage studies such as carcinogenicity, and multi or extended-one-generation and developmental toxicity studies. Work is currently ongoing, and we expect to publish our findings also related to these additional long-term endpoints in the first half of 2026.

Table 10: Proposed add-ons for Smart *in vivo* studies

Tissue/Types of samples	**Possible additional technology**
Biofluids • Blood Plasma • Blood Serum • Urine • Cerebrospinal fluid (CSF)	• Toxicokinetic parameters • Hormonal measurements • Plasma-Metabolomics
Tissues • Fresh • Frozen • Fixed	• Toxicokinetic parameters • Immunohistochemistry (IHC) • Metabolomics/Lipidomics (fresh and frozen only) • Whole genome mRNA transcriptomics • Single gene mRNA measurements e.g. hormones • Gene panel mRNA measurements

Examples of enhanced *in vivo* studies:

> Metabolomics as read-across tool: An example with 3-aminopropanol and 2-aminoethanol - ScienceDirect (Sperber et al. 2019)

The study aimed to use metabolomics to support a read-across approach from 2-aminoethanol (MEA) to 3-aminopropanol (3AP). Wistar rats were treated with MEA and 3AP at different dose levels. Blood samples were collected for metabolome analysis and clinical examinations and body weight measurement were performed. The study demonstrated the high structural and functional similarity between MEA and 3AP, supporting the use of read-across to fill data gaps for 3AP's REACH registration.

> Utility of *in vivo* metabolomics to support read-across for UVCB substances under REACH - PubMed (Kamp et al. 2024)

The study aimed to use metabolomics to support a read-across approach for UVCB substances under REACH. Researchers analysed 15 steam-cracked distillates and six major constituents, including indene and dicyclopentadiene, using plasma metabolomics in a 14-day rat oral gavage study. They identified metabolome patterns and grouped substances based on biological similarity, confirming target organs affected and supporting REACH registration.

> Predicting Drug-Induced Hepatotoxicity Using QSAR and Toxicogenomics Approaches | Chemical Research in Toxicology (Low et al. 2011)

The study aimed to develop hybrid models combining QSAR and toxicogenomics to predict hepatotoxicity. Researchers analysed various chemicals, including acetanilide, thioamide, alkyl chloride, and styrene substructures, using chemical descriptors and toxicogenomics data. They identified toxicophores and improved predictive accuracy for liver toxicity, demonstrating the value of hybrid modelling for identifying chemical alerts.

> PARC Additional Deliverable AD5.2 Guidance on animal studies in PARC, Work Package 5 – T5.1 (PARC 2023)

PARC has the option to conduct and / or subcontract test guideline (TG) conform animal studies to contract research organizations (CRO), if needed, to close regulatory data gaps of concern (task 5.1). To benefit from such studies, on top of the test guideline required endpoints like histopathology or clinical chemistry, additional parameters will be investigated (referred to as "TG+") like toxicokinetics, toxicogenomics of target organs, other data based on alternatives to animal approaches or intra-organ concentrations.

In silico methods

The use of Quantitative Structure–Activity Relationship models ((Q)SAR) tools for predicting toxicological properties within the context of non-animal approaches has become a standard approach.

The software and models commonly used consist of a mix of well-established commercial applications and publicly available, free-to-use, applications that have been extensively employed for regulatory purposes, screening, prioritization, and research. Models include QSAR and SAR based models, and expert systems. Importantly, these models have undergone validation in accordance with the OECD principles (2014) and they are accompanied by QMRF (Quantitative Model Reporting Format) documentation that adheres to guiding principles (OECD Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship (QSAR Models, 2014). Furthermore, the OECD (Q)SAR Assessment Framework (QAF) (2023b; 2024) offers comprehensive guidance for the regulatory application of QSAR models and ultimately contribute to the broader acceptance of (Q)SARs for regulatory purposes (OECD Series on Testing and Assessment No. 386 and No. 405).

While no computation tool has been adopted as OECD test guideline, (Q)SAR can already be applied in various exploratory uses and support regulatory applications. For instance, in adapting information requirements for hazard assessment and classification, the REACH R.6 guideline serves as a foundational reference for employing (Q)SAR approaches.

In practice, (Q)SAR models are more frequently employed within a weight of evidence framework, particularly for assessing more complex endpoints. This approach allows for the integration of various types of information, which can help mitigate the uncertainties associated with (Q)SAR predictions. By combining (Q)SAR data with other relevant information, a more robust assessment can be achieved, ultimately supporting regulatory decision-making.

What would make QSAR predictions more acceptable:

- Development of (Q)SAR models in accordance with regulatory set standards
- Application of models validated according to the OECD guiding principles for (Q)SAR validations (2014)
- Assessment of predictions and reporting in accordance with the requirements set by QAF (2023, 2024).
- Transparent reporting of uncertainties
- Harmonised acceptance criteria across legislations

It would be beneficial if regulators could provide a comprehensive overview of accepted adaptations of standard information requirements based on (Q)SAR predictions, specifying the relevant endpoints. Additionally, incorporating transparent reporting on any rejected adaptations would foster trust and clarity. This approach would not only outline clear directions for improving the use of (Q)SAR predictions in data gap filling but also encourage registrants to submit data generated through (Q)SAR methodologies.

***In vitro* methods**

For the *in vitro* bioactivity assessment, Doe et al. (2025) used publicly available data from the ToxCast database (US EPA 2024) which contains the bioactivity data for over 9,500 chemicals from various

high-throughput screening assays covering a range of endpoints ranging from cell viability and proliferation assays to more specific hormone receptor assays.

The validation status of the assays varies, they are predominantly used for substance screening and prioritization. ToxCast and Tox21 represent a library of medium- and high-throughput *in vitro* assays sponsored by US EPA. In general, each assay assesses the effects of approximately 300-8000 chemicals against a defined target, which ranges from cell culture systems to cell-free assays. Phase I of the ToxCast program focused on a set of around 300 chemicals (primarily agrochemicals), with subsequent phases of the program expanding the list to up to 4000 chemicals. The Tox21 program represents an expanded assessment of 8000 chemicals against a smaller set of targets. Data generated by the ToxCast and Tox21 programs are freely available via the EPA Comptox dashboard as well as via the NTP ICE dashboard.

When using ToxCast data for repeat-dose toxicity assessments, data on both potency (e.g., AC_{50} value) and severity (adverse effects associated with *in vitro* assay target) should be evaluated. For potency, it will be important to consider which point of departure value to select (e.g., AC_{50}, ACC, BMDL) as well as defining potency criteria (i.e., what bioactive concentration ranges indicate high, intermediate and low potency). Moreover, when using any *in vitro* assay to understand *in vivo* hazards, it is important to understand the relevance of the dosimetry. For example, a xenobiotic that inhibits aromatase at high doses (>50 µM) may not be biologically relevant if the maximum attainable internal dose is far lower *in vivo*. As such, *in vitro* to *in vivo* extrapolation (IVIVE) assessments should be undertaken to better understand the likely maximum tolerated dose *in vivo*.

Furthermore, consensus on the severity rating for assay endpoints will be needed to ensure consistency in these assessments. Guidance on how to consider negative assay data also will be needed. When evaluating the use of ToxCast data, it is prudent to understand limitations in assay coverage such as cases where all molecular initiating events within an AOP may not be covered. One example is with skin sensitization where soft electrophilic xenobiotics may be identified in the ARE-NRF2 dissociation assay while no assays to identify hard electrophiles are currently available via the ToxCast database. Moreover, many assays lack metabolism and as such, are unlikely to identify compounds that need to be metabolized to their active toxic form. An example are the phosphothionates, such as malathion, which need to be oxidatively desulphonated to the active oxon in order to inhibit acetylcholinesterase. Furthermore, a more thorough understanding of the MIEs that drive repeat-dose toxicity and a better coverage of these by HTP assays would be needed to more fully predict *in vivo* outcomes. A tiered assessment that incorporates other data streams may help to identify whether there are gaps in biological coverage during specific chemical assessments, an example of this has already been recorded in Doe et al. (2025).

Commercial availability of methods

It is beyond the scope of this document to lay out the commercial availability of all methods referenced. In general, the commercial availability of methods is key to allow method implementation. New method proposals need to be globally available and readily accessible at several CROs to properly cover user demands at sufficient quality. Regulatory certainty with regards to acceptance is needed for CROs to develop a business case for offering a new method.

In the context of the proposed 28-day Smart study design the technologies which are being considered are not new (transcriptomics, metabolomics, measurement of parent and metabolites in blood, specific markers for organ toxicity) and can be implemented with common knowledge and expertise. Therefore, commercial availability will not be a hurdle for its implementation.

As sampling of tissues and fluids (*in vivo*) as well as intra- or extra-cellular matrixes has been shown to be a key component for interlaboratory comparability of results, guidance is needed. Therefore, ECHA has taken the lead of a project to develop the guidance and best practices for sample collection suitable for omics analyses, included under OECD WHPA workplan. This guidance aims to ensure sufficient quality samples, support harmonization through example protocols, provide clarity for all stakeholders on principles for good quality samples and encourage sampling in ongoing studies. The guidance document is being developed by an *ad hoc* drafting group and supported by the OECD Expert Group on Omics. The OECD Omics Reporting Framework (OECD 2023a) is intended to describe the information that should be reported when an omics technology is applied in the context of regulatory decision-making, to enable assessment of the quality of the study from its design through the collection, analysis, and reporting of data.

The guidance, however, does not extend to the interpretation of these data, which is an ongoing ECETOC task force. These activities indicate that these 'omics technologies have reached a state of maturity that would allow for their use in regulatory studies and decision-making processes.

In the context of demonstrating regulatory applications, a case study using the phenoxy herbicides MCPA, Mecprop-p and Dichlorprop-p showed that with the use of a metabolomics in a 28-day study in a read across approach a 90-day study could be waived without affecting the regulatory outcome (van Ravenzwaay et al. 2016). It is acknowledged that proper training of CROs and the development of SOPs will enhance quality and reliability of the obtained results.

For the *in silico* models used by Doe et al., general hazard prediction is well-established, for example for predicting Ames mutagenicity. Additionally, there is a wide range of models available for routinely requested endpoints, such as acute oral toxicity, skin irritation, and eye irritation.

Description of legislative areas that can make use of the approach, timing and recommended steps for regulatory acceptance

The use of a tiered approach and Smart/enhanced *in vivo* studies is possible under the current REACH and CLP legislations but the regulatory acceptance of such an approach is unclear. Protection goals to be met by non-animal approaches need to be defined and this includes the definition of an acceptable protection level for tiered approaches. As laid out in the technical sections of this document, ECETOC believes that current protection goals can be met using Smart *in vivo* studies while allowing for a reduction and refinement of animal testing under the current regulatory paradigms. The implementation of Smart studies should require only limited modification of the current REACH Annex VIII-X (Column 2).

Further, regulatory compliance requires testing proposals for higher tier testing for human health effects to be submitted when developing and refining a REACH registration dossier. As current information requirements are largely based on *in vivo* testing via current OECD test guidelines, there is no time in the registration process to apply a tiered approach to develop a hazard hypothesis and adapt *in vivo* testing protocols towards Smart study design.

Therefore, REACH Annexes on standard information requirements need to be amended to include *in silico* and *in vitro* methods as well as adaptations for Smart study designs. This should not increase the listed methods required to run nor the mandatory requirements but allow flexibility to choose the most suitable approach while still reducing and refining animal testing where possible. REACH Annex XI should be amended to implement the use of tiered approaches similar to the use of Weight of Evidence (Annex XI, Section 1.2) for both demonstrating the presence and absence of hazard and the

classification or non-classification of substances. Specifically for the case of classification, CLP text needs to be amended to lay out the principles for classification based on tiered approaches.

REACH guidance on Information Requirements and Chemical Safety Assessment (ECHA 2025) needs to be amended to lay out the use of tiered approaches for the assessment of systemic toxicity endpoints.

Dialogue between registrants and regulatory agencies is needed in the early implementation phase to refine the use of tiered approaches. Ideally, a new review process for tiered approaches to support ongoing registrations will be implemented.

Regulatory changes to other regulations for biocidal products, food and feed additives, plant protection products, cosmetics, detergents, and medicinal products have been discussed but no conclusive recommendation can be given. It is evident that regulations that include 1:1 dossier reviews and flexibility on how to fulfil information requirements (e.g. cosmetics regulation) are amenable for the adoption of tiered approaches.

Any additional recommendations for research in the area/Further technical work needed

For successful implementation of tiered approaches and enhanced/Smart study designs, guidance is required. Also, additional scientific work can be envisioned to broaden the applicability of Smart studies. The following needs have been identified:

1) Development of Smart studies including reproductive and developmental toxicity endpoints

The initial Smart study project focused on modifications to the 28-day study design as described in OECD TG 407. It is desirable to expand the concept to combined reproductive and developmental toxicity, repeated dose toxicity screening studies (OECD TG 422) as this study type can already be used to fill the Annex VIII information requirement and inform on the design of Annex X studies. Reproductive toxicity studies require special considerations with regards to sampling time points as animals go through sensitive periods like mating, pregnancy and lactation which will have an impact on evaluation, for example for transcriptomics signatures or the metabolome (Henriquez et al. 2024).

2) Further reduction of study duration

Opportunities to further reduce study durations should be explored. Published studies support the use of earlier time points for detection of changes to the transcriptomic signatures or the metabolome (Everett 2024; Sperber et al. 2019). Genotoxicity endpoints also may require a shorter exposure duration only.

3) Develop sampling guidance

A Standard Project Submission Form was approved at the OECD WNT-37 meeting in April 2025 to include an update on sample cryopreservation suitable for omics in OECD TGs 407, 408, 422, 421, 203, 210, 236.

4) Develop guidance documents to lay out the experimental design of enhanced/SMART *in vivo* studies

Including but not limited to dose level setting, sampling time points and number of animals per dose group.

5) Develop or adapt guidance on data analysis and reporting

ECETOC has been actively involved in advancing the use of omics technologies in regulatory toxicology. The Regulatory Omics Data Analysis Framework (R-ODAF) provides guidelines for analysing omics data for regulatory purposes and helps in consistent and accurate interpretation of complex omics datasets (Verheijen et al. (2020); Verheijen et al. (2022)). Omics Data Reporting and Documentation (OORD) focuses on establishing standards for reporting and documenting omics data to ensure data used in regulatory assessments are of high quality and reproducible. Work is ongoing to develop an Omics Data Interpretation Framework for Regulatory Application (ODIFRA) (Gant et al. 2023). These efforts will support the implementation of Smart studies.

6) Develop case studies to exemplify the use of enhanced studies to demonstrate biological similarity in read-across

Enhanced studies have shown promise for the demonstration of biological similarity and to support grouping and read-across (OECD GD194, under revision). Further case studies should be conducted and published to support the evidence. Similarly, case studies exemplifying the use of Smart studies to waive higher tier testing should be published.

7) Further ECETOC work

ECETOC foresees to conduct and publish a fully integrated case study on the Staged Assessment Framework (*in silico*/*in vitro*/Smart *in vivo* tiers).

References

Ball N, Bars R, Botham PA, et al. (2022) A framework for chemical safety assessment incorporating new approach methodologies within REACH. Archives of Toxicology 96(3):743-766 doi:https://doi.org/10.1007/s00204-021-03215-9

Chapman K, Josephine B, Marco C, et al. (2014) Reducing Pre-Clinical Blood Volumes for Toxicokinetics: Toxicologists, Pathologists and Bioanalysts Unite. Bioanalysis 6(22):2965-2968 doi:https://doi.org/10.4155/bio.14.204

Doe JE, Botham P, Holland D, et al. (2025) Framework for classifying chemicals for repeat dose toxicity using NAMs. Archives of Toxicology doi:10.1007/s00204-025-04069-1

ECHA (2025) Guidance on Information Requirements and Chemical Safety Assessment. In. https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment Accessed May 2025

EPAA (2023) European Partnership for Alternative Approaches to Animal Testing launches Designathon for Human Systemic Toxicity'. European Commission. In. https://single-market-economy.ec.europa.eu/calls-expression-interest/epaa-launches-designathon-human-systemic-toxicity_en Accessed March 2024

Everett L, L. Wehmas, K. Vitense, S. Davidson-Fritz, J. Lambert, J. Cowden, M. Hughes, A. Harrill, S. Auerbach, M. Devito, R. Thomas (2024) Deriving transcriptomic reference values for chemicals from 5-day repeat dose exposure studies in rodents. 55th Annual Meeting of the Environmental Mutagenesis and Genomics Society, Palm Springs, CA, September 07 - 11, 2024 doi:https://doi.org/10.23645/epacomptox.27068329

Gant TW, Auerbach SS, Von Bergen M, et al. (2023) Applying genomics in regulatory toxicology: a report of the ECETOC workshop on omics threshold on non-adversity. Archives of Toxicology 97(8):2291-2302 doi:https://doi.org/10.1007/s00204-023-03522-3

Henriquez JE, Badwaik VD, Bianchi E, et al. (2024) From Pipeline to Plant Protection Products: Using New Approach Methodologies (NAMs) in Agrochemical Safety Assessment. Journal of Agricultural and Food Chemistry 72(19):10710-10724 doi:https://doi.org/10.1021/acs.jafc.4c00958

Kamp H, Kocabas NA, Faulhammer F, et al. (2024) Utility of in vivo metabolomics to support read-across for UVCB substances under REACH. Archives of Toxicology 98(3):755-768 doi:https://doi.org/10.1007/s00204-023-03638-6

Low Y, Uehara T, Minowa Y, et al. (2011) Predicting Drug-Induced Hepatotoxicity Using QSAR and Toxicogenomics Approaches. Chemical Research in Toxicology 24(8):1251-1262 doi:https://doi.org/10.1021/tx200148a

OECD (2023a) OECD Omics Reporting Framework (OORF): Guidance on reporting elements for the regulatory use of omics data from laboratory-based toxicology studies. In: OECD Series on Testing and Assessment, No 390, OECD Publishing, Paris. https://doi.org/10.1787/6bb2e6ce-en

OECD (2023b) (Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions. In: OECD Series on Testing and Assessment, No 386, OECD Publishing, Paris. https://doi.org/10.1787/d96118f6-en

OECD (2024) (Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions, Second Edition. In: OECD Series on Testing and Assessment, No 405, OECD Publishing, Paris. https://doi.org/10.1787/bbdac345-en

PARC (2023) PARC Additional Deliverable AD5.2 Guidance on animal studies in PARC, Work Package 5 – T5.1. In. https://www.eu-parc.eu/sites/default/files/2023-08/PARC_AD5.2.pdf Accessed May 2025

Sperber S, Wahl M, Berger F, et al. (2019) Metabolomics as read-across tool: An example with 3-aminopropanol and 2-aminoethanol. Regulatory Toxicology and Pharmacology 108:104442 doi:https://doi.org/10.1016/j.yrtph.2019.104442

USEPA (2024) United States Environmental Protection Agency. Exploring ToxCast Data. Toxicity Forecasting In. https://www.epa.gov/comptox-tools/exploring-toxcast-data Accessed March 2024

van Ravenzwaay B, Sperber S, Lemke O, et al. (2016) Metabolomics as read-across tool: A case study with phenoxy herbicides. Regulatory Toxicology and Pharmacology 81:288-304 doi:https://doi.org/10.1016/j.yrtph.2016.09.013

Verheijen M, Tong W, Shi L, Gant TW, Seligman B, Caiment F (2020) Towards the development of an omics data analysis framework. Regulatory Toxicology and Pharmacology 112:104621 doi:https://doi.org/10.1016/j.yrtph.2020.104621

Verheijen MCT, Meier MJ, Asensio JO, et al. (2022) R-ODAF: Omics data analysis framework for regulatory application. Regulatory Toxicology and Pharmacology 131:105143 doi:https://doi.org/10.1016/j.yrtph.2022.105143

Worth AP, Berggren E, Prieto P (2025) Chemicals 2.0 and Why We Need to Bypass the Gold Standard in Regulatory Toxicology. Altern Lab Anim 53(1):21-25 doi:https://doi.org/10.1177/02611929241296328

4.5. Developmental neurotoxicity - the DNT in vitro battery

This chapter was originally developed by one of the working group members (Swiss Centre for Applied Human Toxicology, SCAHT) for the human health working group, which reviewed its content before inclusion into the SWD.

Implementing change for developmental neurotoxicity (DNT) assessments

Summary

Developmental neurotoxicity (DNT) testing is undergoing a paradigm shift from animal-based studies to mechanistic, human-relevant alternatives to animal testing. Traditional *in vivo* assays are resource-intensive, slow, and of limited translational value, leaving most chemicals untested for potential effects on brain development. The DNT In Vitro Battery (DNT IVB) addresses these limitations through 17 complementary human-relevant assays covering key neurodevelopmental processes such as neural proliferation, migration, differentiation, and synaptogenesis.

Over the past 20 years, considerable advancements have been made to develop assays measuring KNPs. The early scientific workshops and the regulatory necessity led to a movement to implement NAMs into chemical DNT assessment, by EFSA, and US EPA, overseen by OECD. This multiyear effort provided the foundation for an OECD project that started in 2018, which described DNT assays, provided criteria that allow evaluation of the relevance of the data to DNT, and helped determine the degree of certainty of the results to inform the use of a DNT in vitro battery (DNT-IVB) in regulatory hazard determinations. In April of 2023, the OECD Working Party of National Coordinators of the Test Guideline program approved document 377 that described the DNT-IVB (OECD, 2023).

Currently multiple coordinated initiatives with several synergistic projects are underway to accelerate the development, standardization and regulatory use of NAMs for neurotoxicity. Within the European Union, these include large collaborative efforts such as PARC, ONTOX, RiskHunt3r, CHIASMA, EFSA Environmental Neurotoxicants activities, and EFSA funded Brain Health project, (Tal et al., 2024; EFSA., 2025). In parallel, there are various initiatives led by NIEHS, FDA and US EPA. OECD continues to support coordination and capacity building through guidance documents, webinars, and training materials, facilitating consistent interpretation and uptake across regulatory jurisdictions. As such, the DNT IVB can be used for hazard and risk assessment. A tiered testing framework is proposed, integrating computational (Tier 0), in vitro (Tier 1), confirmatory mechanistic (Tier 2), and complex model (Tier 3) approaches, enabling both hazard and risk assessment applications. Implementation for regulatory use requires legislative alignment and structured change management across sectors, guided by short-, medium-, and long-term actions encompassing policy adaptation, scientific refinement, best practices, and communication.

The roadmap discusses a short-medium and long-term vision for a more globally harmonized, animal-free strategy for DNT assessment that ensures high levels of health protection while advancing scientific innovation and regulatory efficiency. For a recent OECD review, please consult the recent publication "From Concept to Regulation: OECD's Role in the development of the DNT In Vitro Battery" (Sachana et al., 2026) [34].

History of the approach: why a change for DNT assessments?

Prevalences of neurodevelopmental disorders like autism spectrum disorder, attention deficit hyperactivity disorder or IQ loss have been on an international rise [1, 2, 3, 4, 5]. Such diseases pose an immense burden to individuals, families and society [6]. While differences in diagnosis have been identified as contributors to culmination of neurodevelopmental diseases, also specific chemicals have been strongly associated with adverse neurodevelopmental outcome in humans [7, 8, 9, 10]. Regulatory evaluation of the potential of compounds to alter the structure or function of the nervous system during pre- and/or postnatal life (developmental neurotoxicity, DNT) has traditionally depended on *in vivo*

animal testing. Guideline studies conducted under U.S. EPA or OECD protocols are expensive, lengthy, and largely restricted to a narrow set of apical behavioural endpoints. Their interpretation is further complicated by methodological limitations and, as with many animal-based approaches, substantial uncertainty when extrapolating results to humans [11]. Consequently, **thousands of chemicals remain insufficiently assessed for their potential to interfere with nervous system development, leaving considerable gaps in hazard evaluation**. To overcome these shortcomings, new approach methodologies (NAMs) are being advanced to generate mechanism-based and human-relevant information on chemical hazards and risks while reducing reliance on animal use. Alternatives to animal testing integrate multiple complementary tools, including human-derived in vitro systems (e.g., neural cell cultures), alternative species models (e.g., zebrafish), advanced chemical analytics, and computational in silico approaches. Over the past decade, coordinated international efforts have fostered the development and application of the DNT In Vitro Battery (IVB), a large panel of assays targeting key neurodevelopmental processes such as neural proliferation, migration, differentiation and neural network formation. The DNT IVB is intended to provide regulatory authorities with mechanistic data suitable for supporting chemical DNT hazard assessments [12].

The DNT IVB

The DNT IVB was set up to cover a broad variety of key neurodevelopmental processes (Figure 12), which are fundamental for brain development.

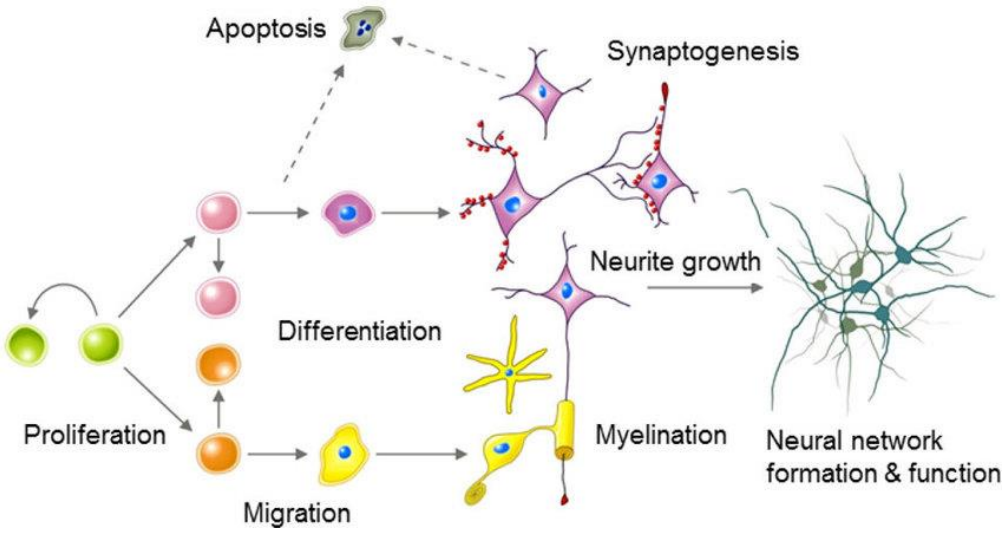


Figure 12: Schematic representation of key neurodevelopmental processes. These processes cover the basics of brain development. Figure from [13].

Seventeen assays currently cover this depicted biological applicability domain (Table 11). The assays have been intensively characterized and described [14] and also used for mode-of-action analyses, screening and AOP development (respective references can be found in [15, 12]).

Table 11: Assays currently in the DNT In Vitro Battery [12].

Test Method (Assay)	Test System (Cell culture)	Assay Duration/ Chem exposure	DNT Endpoint	Viability/Cytotoxicity Endpoint
Proliferation				
NPC1	human NPC grown as proliferating 3D neurospheres	72 h / 72 h	neurosphere area, BrdU incorporation in dividing cells	Resazurin reduction /LDH release

hNP1 Prolif	human NPC	24 h / 24 h	BrdU incorporation in dividing cells	ATP level
Apoptosis				
hNP1 Apop	human NPC	24 h / 24 h	apoptosis pathway (Caspase) activation	ATP level
Migration				
UKN2	human NSC-derived neural crest cells	72 h / 24 h	number of cells moving into defined area	Calcein-AM vital dye
NPC2a	human NPC grown as differentiated 3D neurospheres	72 h / 72 h 120 h / 120 h	mean distance of radial glia (nuclei negative for neuronal and oligodendrocyte markers) from edge of sphere	Resazurin reduction/LDH release
NPC2b	human NPC grown as differentiated 3D neurospheres	120 h / 120 h	mean distance of tubulin-positive neurons from edge of sphere	Resazurin reduction/LDH release
NPC2c	human NPC grown as differentiated 3D neurospheres	120 h / 120 h	mean distance of O4-positive oligodendrocytes from edge of sphere	Resazurin reduction/LDH release
Neuronal Differentiation				
NPC3	human NPC grown as differentiated 3D neurospheres	120 h / 120 h	number of tubulin-positive neurons	Resazurin reduction/LDH release
Neurite outgrowth				
NPC4	human NPC grown as differentiated 3D neurospheres	120 h / 120 h	neurite length & area	Resazurin reduction /LDH release
UKN4	human NSC-line (v-myc transformed)	72 h / 24 h	neurite area	Calcein-AM vital dye
UKN5	human iPSC-derived peripheral (sensory) neurons	24 h / 24 h	neurite area	Calcein-AM vital dye
hN initiation	human NPC-derived neurons	48 h / 48 h	neurite length	cell morphology
Cortical initiation	rat primary neocortex	48 h / 48 h	neurite length	cell morphology
Neurite Maturation and Synaptogenesis				
Cortical maturation	rat primary neocortex	288 h / 120 h	dendrite length	cell morphology
Cortical synapto	rat primary neocortex	288 h / 120 h	synapse number	cell morphology
Glial Differentiation				
NPC5	human NPC grown as differentiated 3D neurospheres	120 h / 120 h	number of O4-positive oligodendrocytes	Resazurin reduction /LDH release
Neural Network Formation				
Cortical MEA	rat primary neocortex	288 h / 288 h	Action potential spike and burst parameters related to network connectivity	Resazurin reduction/total LDH

As the key neurodevelopmental processes covered by the DNT IVB are cellular key events (KE), a generic AOP was suggested with the cellular responses representing the endpoints covered by the battery (Figure 13).

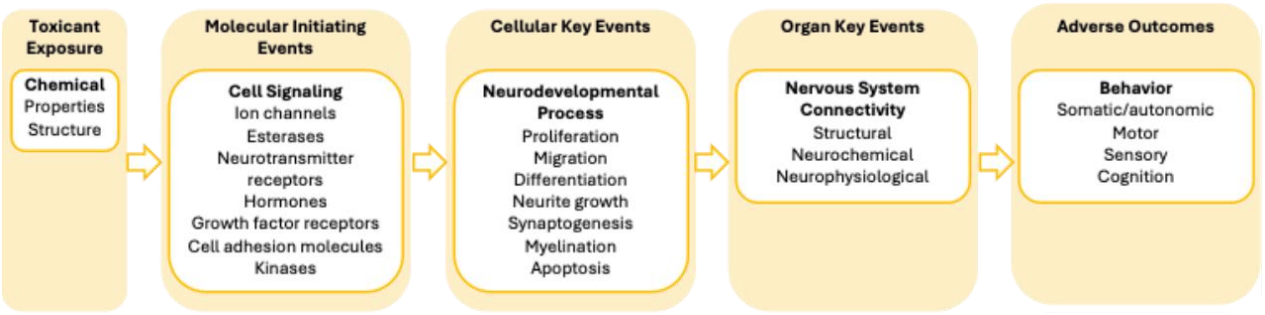


Figure 13: Diagram illustrating a selection of components of a generic Adverse Outcome Pathway (gAOP) for DNT. The gAOP shows the relationship between MIEs such as altered cell signaling, and key events (KEs) at the cellular level (neurodevelopmental processes) and organ level (nervous system connectivity), which lead to an adverse outcome (AO) for the individual. Modified from [12].

Mapping Initiatives and Securing the Learnings

The DNT IVB was assembled in 2018 [16, 17] according to established readiness criteria [18]. Initially developed by two laboratories in the EU and one in the US, the battery was initially challenged with positive and negative compounds proving the test methods' performance not only for mode-of-action analysis of few compounds, but also as screening tools with broad chemical and biological applicability domains [19, 20]. During the process of OECD recommendation, OECD member states and international agencies commented on the use, reporting and performance of the DNT IVB. While ready to be used immediately [12], more efforts on lab-to-lab transfer, glia cell-mediated DNT, IATA (integrated approaches to testing and assessment) case studies, adverse outcome pathways (AOPs) as well as training were thought to benefit the overall trust in the DNT IVB. To accommodate these matters, there are a variety of initiatives on the way to support the DNT IVB on its path to regulatory application:

- EFSA Working Group on quantitative in vitro to *in vivo* extrapolation (QIVIVE) of the DNT IVB data (2024 - 2026).
- The EFSA DNT-RAP2 project is transferring all 17 assays (Table 11) of the DNT IVB from the developers` laboratories to an independent laboratory of a contract research organization (www.dntox.de; 2024 - 2027).
- The US Division of Translational Toxicology at the National Institute of Environmental Health Sciences (NIEHS) has assigned testing of >300 compounds in most assays of the DNT IVB including zebrafish (https://www.niehs.nih.gov/research/atniehs/dtt/strategic-plan/health/developmental). While data of the first +100 compounds are already available in ToxCast, the second +100 dataset is produced, but not published yet and the third +100 substances are currently being tested (2022 - 2026).
- The EFSA Brain Health project (GP/EFSA/ED/2022/01) aims at understanding the current coverage of glia-related processes in the DNT IVB by focusing on astrocytes, microglia and myelinating oligodendrocytes. Single cell RNA Seq will inform on molecular/cellular aspects. Novel glia-focused AOPs are being set up (2023 - 2027).

- EFSA assigned testing of +100 pesticides (substances with positive and negative outcomes in the DNT *in vivo* guideline study and compounds with no available DNT *in vivo* data) in the 17 assays of the DNT IVB (DNT-RAP1; 2024 - 2027). Results will allow drawing a conclusion of the DNT IVB to be suitable for compound classification within the CLP framework.
- The ASPIS cluster project ONTOX (https://ontox-project.eu) is creating ontologies for DNT based on physiological maps and AOP networks (2021 - 2026).
- The ASPIS cluster project RiskHun3r (https://www.risk-hunt3r.eu) produces DNT IATA case studies (2021 - 2026).
- The public-public Partnership for the Assessment of Risk from Chemicals (PARC) is involved in several DNT activities. Refinement of DNT IVB assays, AOP development [21] and DNT regulatory case studies including application in a possible new hazard class for neurotoxicity as an ECHA initiative (https://youtu.be/Ts55QFMnor4; 2022 - 2029).
- The EFSA funded DNT-RAP4 project will provide four IATA case studies on selected compounds that diverge between DNT *in vivo* and in vitro results (2024 - 2027).
- The CHIASMA project will provide solutions for DNT and other alternatives to animal testing applications in CLP regulation under REACH (2024 - 2027).
- The OECD is publishing a series of webinars as training material for the DNT IVB (https://www.youtube.com/playlist?list=PLJNHHjqEVIIfn6eDbh1N-tLAhxP45zBft; 2025-2026).

These activities will further reduce the known main uncertainties, which are highlighted in the OECD initial recommendations from 2023 [12]. These include general uncertainties not specific for the DNT IVB and DNT IVB-specific uncertainties. Initiatives addressing respective uncertainties are given (*in brackets in italics*):

1. General uncertainties concerning in vitro testing:
    - Toxicokinetics. ADME (absorption, distribution, metabolism, and excretion) of the DNT IVB assays are limited towards developing brain cells and do not contain e.g. maternal and fetal liver metabolism or barriers like the placenta or the developing blood-brain-barrier (*EFSA QIVIVE Working Group*).
    - Endocrine/immune systems. A variety of systemic impacts of hormones and immune signaling on brain development are not captured in the current DNT IVB (*EFSA BrainHealth and PARC projects*).
    - Chemical domains. So far, there is a relatively limited number of tested compounds in the DNT IVB. Certain compound classes (for example volatile compounds) remain challenging in vitro (*EFSA, NIEHS and PARC chemical testing*).
    - Inter- and intra-species differences. Some assays still rely on animal-derived neural cell cultures. The human cell-systems have no full coverage of sex or human genetic diversity (*PARC & EFSA Brain Health projects*).

2. DNT IVB-specific uncertainties:
    - Currently incomplete coverage of cell-to-cell interactions in some test systems (*EFSA Brain Health project*).
    - Limited availability of DNT AOPs (*EFSA Brain Health, PARC, ONTOX and CHIASMA projects, SCAHT AOP_HUB*).
    - Negatives/Sensitivity: False negative findings may occur due to a current lack of coverage of critical neurodevelopmental processes (*EFSA Brain Health project, EFSA RAP4 project*).
    - Positives/Specificity: False positive results may occur based on the uncertainties listed above, mainly kinetics (e.g. no detoxification kinetics, no placental and blood-brain-barrier

modelled) or single cell type representation in some of the assays (e.g. neurons with lack of astrocytes) (*EFSA QIVIVE Working Group, EFSA Brain Health project, EFSA RAP4 project*).

- Validation status/Inter-laboratory testing: While no formal OECD Guidance Document 34 validation procedure was conducted, DNT IVB assays were scientifically validated for relevance and robustness. An inter-laboratory testing and transfer of the DNT IVB is needed (*EFSA DNT-RAP2 project*).

While the current DNT IVB is ready-to-use, it will further evolve with scientific innovation that will feed into the lifecycle of DNT NAMs.

This fits into the context of a six-step framework of the DNT IVB, from basic research to regulatory application (Figure 14). It serves as a case example [22] that might be utilized by other fields, e.g. Developmental and Reproductive Toxicity (DART) [23]. The framework describes how the DNT IVB assays went from development through key steps towards regulatory acceptance. Scientific validation was core during this process ensuring biological relevance and description of applicability domain of the assays, toxicological domain covered and intra-laboratory reproducibility.

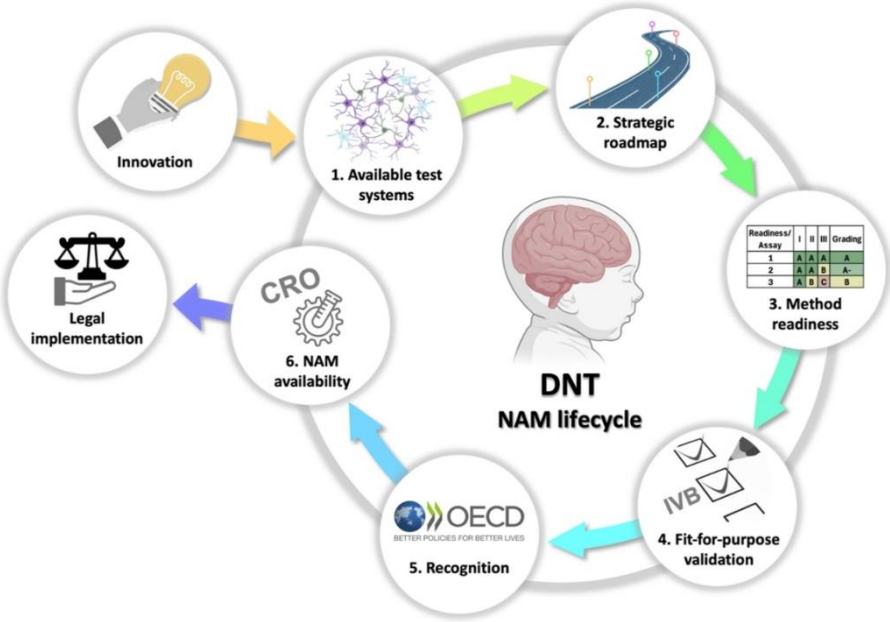


Figure 14:The DNT New Approach Methodology (NAM) lifecycle: a six-step framework to facilitate regulatory implementation. The DNT New-Approach Methodology lifecycle is divided into 6 key steps, that reflect the development of the DNT IVB from scientific innovation towards readiness for legal implementation. Figure from [22].

An ongoing step is the inter-laboratory transfer to a naive laboratory (*EFSA RAP-2*, see listed initiatives and uncertainties above). In making the DNT-IVB publicly available, this critical step will further facilitate the application of the DNT IVB in regulatory frameworks. For this, the DNT IVB could be embedded in a defined tiered testing framework or a future defined approach.

Proposed DNT testing Framework

The OECD Expert Group on DNT in vitro has been guiding the establishment, evaluation and regulatory use of the DNT IVB over many years. The OECD initial recommendations were a major step on the DNT IVB road to regulatory acceptance and application.

One possible use of the DNT IVB is placing it into a framework like an IATA [24] or a defined approach (DA). The framework might be flexible depending on the regulatory question and is currently under discussion at the OECD. There has been the unanimous agreement that the DNT IVB is one step of a potential tiered approach and does not contain tiers in itself. This agreement is based on the biological rationale of the battery: each key neurodevelopmental process represented by a battery assay is equally important, i.e. disturbances in proliferation of neural progenitor cells or alteration of differentiation or synaptogenesis will lead to an adverse outcome (Figure 13). Chemicals might affect different endpoints of the DNT IVB with different potencies rectifying the current need for running all assays of the battery for capturing the most sensitive endpoint as the point of departure (PoD) for risk assessment.

Use of the DNT IVB entails two major applications, (i) hazard assessment for e.g. screening and prioritization and (ii) risk assessment.

(i) For hazard assessment purposes the battery is ready to be used and needs to be applied in its totality of the 17 assays. If available, an in silico QSAR approach might precede the in vitro testing, however at this point in time QSAR will not be a stand-alone approach for DNT hazard assessment. In addition, considerations of the impact of metabolism and the bioactivity of metabolites within the DNT IVB needs to be considered. After receiving benchmark concentrations (BMC) from the 17 assays, the most sensitive endpoint(s) can be identified. The most sensitive BMC can be used for calculating the effective from the nominal in vitro bioactive compound concentration. Such an approach will solely inform on hazard for studied key neurodevelopmental processes and could be followed up by a second hit assay. This might be advisable in case the hit occurred in an assay with sole neuronal cultures lacking glia or in rat cell-based assays.

(ii) For risk assessment purposes a tiered approach around the DNT IVB is currently developed and a first outline of such an approach was recently discussed in the OECD `Workshop on critical innovations in pesticides safety testing and chemical risk assessment for developmental neurotoxicity (DNT)` [25] following previous considerations [26] resulting in a published proposal in the workshop proceedings [27]. This tiered testing proposal is summarized in the next paragraph. However, there are various uncertainties in this proposal, among them, (1) a systematic analysis of all DNT available methods and sources of information has not been done; (2) the predictive capacity of each of the tiers has not been defined; (3) for some of the tiers readiness criteria for regulatory use have not been discussed or demonstrated (e.g., tier 1 TK, QSAR models or *in vivo* targeted studies); (4) the entire workflow in this strategy has never been employed for any case study.

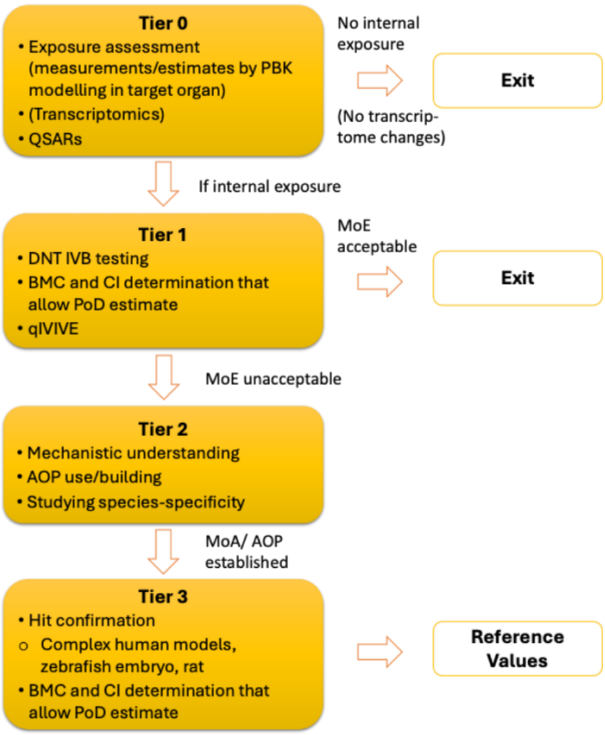


Figure 15: A proposed tiered testing approach for DNT as discussed at the 2024 OECD workshop on critical innovations in pesticides safety testing and chemical risk assessment for developmental neurotoxicity (DNT) [25, 27].

The proposed tiered strategy consists of four tiers and aligns with the IATA concept (Figure 15).

Tier 0 strongly relies on computational tools and modelling data (e.g., QSAR models, physiologically-based kinetic (PBK) modelling) to predict DNT potential and model human exposure levels. An acceptable margin of exposure might lead to an exit at Tier 0, whereas internal exposure or hazard alerts from Tier 0 will guide into Tier 1.

Tier 1 then applies the DNT IVB [12, 19, 20,], running all 17 assays in parallel to identify effects on key neurodevelopmental processes and to derive benchmark concentrations. Importantly, these nominal concentrations are translated to *in vivo* equivalent doses using quantitative in vitro to *in vivo* extrapolation (qIVIVE). If a derived margin of exposure is considered safe at this level, there is an exit point at Tier 1.

Tier 2 includes confirmatory or orthogonal assays to address remaining uncertainties or refine mechanistic understanding. This can be achieved via available or by building new AOPs. Another approach is to include mechanism-based high-throughput assays [28], cell painting assays [29, 30] or by performing (high throughput) transcriptomics [31].

Tier 3 is thought to be optional, yet preferred, for hit confirmation. For this, more complex model systems like 3D cultures, human BrainSpheres, or organoids can be utilized. At this stage, also immunological competence may be implemented in BrainSpheres, for example, by adding human induced pluripotent stem cell-derived microglia. At Tier 3, testing in the zebrafish embryo as an alternative to mammalian testing may be applied. Its main advantages are biological complexity and early behavioral readouts [21]. However, species differences, mechanistic relevance, and predictive value still need further investigation. Also, while considered a non-animal model in the EU, it remains a vertebrate, raising some ethical debate. Alternatively, targeted or as a very last resort DNT TG *in vivo* or behavioural tests using the rat might be performed as a Tier 3 confirmatory testing.

Change Management Requirements for Implementation

The case of DNT differs from other testing endpoints concerning the implementation of non-animal approaches as in Union legislation, testing of chemicals for their adverse DNT effects is not mandatory, yet has to be performed under some regulations upon triggers of adult neurotoxicity or EATS (estrogen, androgen, thyroid, steroidogenesis)-mediated endocrine disruption. This legislative void has left a very large number substances untested for DNT. Therefore, effective change management for DNT hinges on two levers, which open the same door: (i) implementing mandatory DNT evaluation of chemicals into legislation across sectors and (ii) securing regulatory acceptance of the DNT in vitro battery as the only way how testing of large numbers of chemicals can be achieved.

Science-based considerations have been the basis for a new approach to DNT assessment since the beginning of the journey [16, 17]. By forming the International STakeholder NETwork (ISTNET) for DNT, a testing roadmap as change management was agreed on by necessary stakeholders at that time: scientists from academia, international regulatory agencies and industry [32]. This early multi-stakeholder effort was followed by placing the evolution of DNT testing using alternatives to animal testing under the OECD umbrella [33], which has led to the initial recommendations of the OECD [12] (phase 1 commitment). Now the phase 2 commitment needs further management and buy-in (Figure 16). Relevant actions need definition towards supporting the implementation of the new approach for DNT assessment:

- Communication: Policy makers, regulators, public to align on need for more DNT assessments across regulations and where useful as well as means to conduct them.
- Policy roadmap: step-by-step implementation for plant protection products, biocidal products, REACH-chemicals, other chemicals in scope of the roadmap like UV filters, colorants, food and feed additives.
- Science-based initiatives: Harmonization of efforts towards the common goal
- Best practices: IATA, DA, AOP

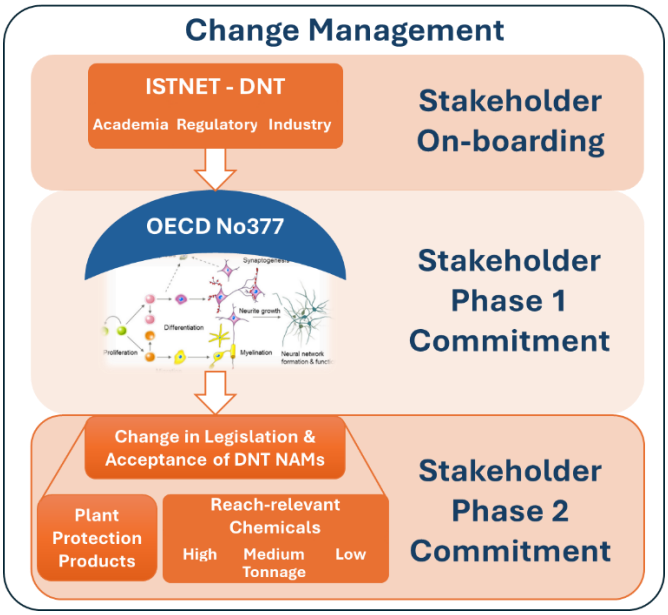


Figure 16: Change management for implementation of the DNT IVB. Stakeholder on-boarding towards changes in regulation for DNT took place early [32]. Phase 1 commitment was achieved by the OECD initial recommendations for the DNT IVB [12]. Now the Phase 2 commitment phase will lead to changes in legislation and the acceptance of the DNT IVB for regulatory application.

Summary Action Plan for Implementation of Framework

Presented below are the key actions for the short, medium and long-term, presenting requirements for a phased transition towards implementation of alternatives to animal testing for DNT assessment.

Short-Term Actions – Use of the DNT IVB for the assessment of active substances use in plant protection products *Policy*

- Strongly prioritize DNT testing due to its fundamental societal significance.
- Achieve a revision of the data requirements for active substances used in plant protection products, introducing the option to consider use of the DNT-IVB where an assessment of DNT is required.
- Secure funding for case studies executing the proposed tiered testing approach.
- Establish safe spaces to stimulate dialogue between industry and regulators on data assessment.

Science

- Improve computational methods for qIVIVE and hazard identification.
- (Further) develop and validate more complex alternatives to animal testing, like brain organoids and model organisms, for higher tier testing as a hit confirmation of tier 1 DNT IVB hazard assessment (see above) not using animals.
- Establish and test WoE-based IATAs and DA in a tiered testing format with the DNT IVB as a central tool.
- Continue refining the DNT IVB according to test method redundancy, human-specificity and costs as commenced within the PARC project.
- Perform uncertainty assessment of alternative methods-based approach for DNT assessment and compare to traditional approach and define level of protection.

Best Practices

- Establish a coordinated platform for collaboration between related projects/initiatives, ensuring synergy and co-creation.
- Develop guidance for WoE-based approaches for DNT assessment, based on e.g. lessons learned from (OECD) IATA case studies.
- Explore together with ECHA and JRC the possibility of implementing a new hazard class for neurotoxicity (including DNT).
- Support availability of DNT alternatives to animal testing by contract research organizations.

Communication

- Develop a communication roadmap encompassing the needs identified for policy roadmap, science and best practices, differentiated where needed.
- Define and communicate benefits of the new paradigm especially concerning protection goals for consumers.
- Continue international stakeholder exchange to ensure alignment and success.
- Organize discussion and consultation fora.

Medium-Term Actions – Consolidation and Confidence Achieved

Policy

- Assess and align for which legislations DNT testing should be mandatory.

- Adapt the regulatory processes and legislations for REACH-chemicals that enable a WoE alternative method-based approach for DNT assessment (possibly tiered by tonnage).
- Explore adapting CLP/GHS criteria for neurotoxicity including DNT alternatives to animal testing.
- Ensure dialogue between submitters and assessors via safe spaces during the assessment process.
- Secure funding for the development of missing DNT-IVB alternatives to animal testing as well as for translational activities.
- Monitor and review 3Rs impact of changed approach for DNT assessment; adapt long(er)-term goals of the roadmap where needed.

Science

- Include scientific progress on alternatives to animal testing into the DNT IVB (e.g. human neurodevelopmental diseases, immune system function, sensitive windows of exposure like puberty or old age)
- (Further) Develop guidance for WoE-based approach for DNT assessment
- Improve and refine internal TTC approach.
- Conduct read-across case studies on the basis of DNT in vitro alternatives to animal testing, designed in co-creation with stakeholders, to increase the acceptance of read-across for carcinogenicity assessment.

Best Practices

- Utilize the revised OECD GD34 for validation for the DNT alternatives to animal testing.
- Training for all relevant stakeholders, including application of WoE-based approach and hands-on lab training for validation and onboarding of new alternatives to animal testing.

Communication

- Prepare materials and deliver outreach, communication and training on the WoE-based approach for DNT assessment.

Long-Term Actions – Non-Animal Approach Implemented

Policy

- Secure funds for applying best science in regulatory toxicology (translation of basic science methods into use for regulation).

Science

- Ensure that the DNT alternative method-based approach follows the evolvement of science. Make use of best sciences.

Best Practices

- Use sector-specific cases to optimize data interpretation and methodologies.
- Ensure all global guidance documents are aligned with one approach for DNT assessment.
- Review progress and learnings of newly implemented animal-free approach for DNT assessment.
- Implement training of next generation professionals (including students, regulators, industry) on new approaches for DNT assessment.

Communication

- Communicate success.

Conclusion

The phases and completion of the actions summarized here rely on a continuum of progress from short-mid-long-term achievable for policy, science, best practices and communication. The collective goal is to implement the animal-free approach successfully for DNT assessment globally across all sectors to reach the highest protection level for the developing brains.

References

[1] UK: Russell et al., 2022. doi: 10.1111/jcpp.13505
[2] DK: Jensen and Steinhausen 2015. doi: 10.4088/JCP.14m09094
[3] US: Dvorak et al. 2023. doi: 10.1016/j.intell.2023.101734
[4] Global meta-analysis: Talantseva et al., 2023. doi: 10.3389/fpsyt.2023.1071181
[5] US: https://www. cdc.gov/mmwr/volumes/74/ss/ss7402a1.htm
[6] EU: Bellanger et al., 2015. doi: 10.1210/jc.2014-4323
[7] Grandjean and Landrigan 2014. doi: 10.1016/S1474-4422(13)70278-3
[8] Miotto et al., 2024. doi: 10.1080/09603123.2024.2384963
[9] EFSA 2019. doi: 10.2903/j.efsa.2019.5809
[10] Bessems et al., 2023. Eionet Report – ETC HE 2023/14
[11] Paparella et al., 2020. doi: 10.1016/j.reprotox.2020.08.002
[12] OECD No 377, 2023. https://doi.org/10.1787/91964ef3-en
[13] Aschner et al., 2017. doi: 10.14573/altex.1604201
[14] Annex B of OECD 377, 2023. https://doi.org/10.1787/91964ef3-en
[15] Crofton and Mundy 2021. https://doi.org/10.2903/sp.efsa.2021.EN-6924
[16] Lein et al., 2005. doi: 10.1016/j.etap.2004.12.035
[17] Fritsche et al., 2018. doi: 10.1093/toxsci/kfy175
[18] Bal-Price et al., 2018. doi:10.14573/altex.1712081
[19] Carstens et al., 2022. https://doi.org/10.1093/toxsci/kfac018
[20] Blum et al., 2023. https://doi.org/10.1016/j.chemosphere.2022.137035
[21] Tal et al., 2024. doi: 10.3389/ftox.2024.1359507
[22] Blum et al., 2025. doi: 10.1016/j.cotox.2025.100528
[23] Fritsche et al., 2024. doi:10.14573/altex.2410081
[24] Hernandez et al., 2021. doi: 10.2903/j.efsa.2021.6599
[25] OECD Workshop 2024. https://www.oecd.org/en/events/2024/10/workshop-critical-innovations-in-pesticides-safety-testing-and-chemical-risk-assessment-for-developmental-neurotoxicity.html
[26] Masjosthusmann et al., 2020. https://doi.org/10.2903/sp.efsa.2020.EN-1938
[27] OECD Workshop Proceedings 2025. https://one.oecd.org/document/ENV/CBC/MONO(2025)10/en/pdf
[28] Jeong et al., 2022. doi: 10.1016/j.tiv.2022.105451
[29] Cimini et al., 2023. doi: 10.1038/s41596-023-00840-9
[30] Nyffeler et al., 2023. doi: 10.1016/j.taap.2023.116513.
[31] Harrill et al., 2024. doi: 10.1093/toxsci/kfae108.
[32] Bal-Price et al., 2015. doi: 10.1007/s00204-015-1464-2
[33] Fritsche et al., 2018. doi: 10.1016/j.taap.2018.02.004
[34] Sachana M et al., Frontiers in Toxicology, 2026 accepted in press

4.6. Developmental and reproductive toxicity (DART) assessments

This chapter was originally developed by one of the working group members (Swiss Centre for Applied Human Toxicology, SCAHT) for the human health working group, which reviewed its content before inclusion into the SWD.

Summary

Developmental and Reproductive Toxicity (DART) assessment currently depends largely on animal studies and is among the most animal-intensive domains of regulatory toxicology (Rovida & Hartung, 2009). These studies are ethically debated, time- and resource-intensive, and their translation to humans is limited by inter-species differences. Important human-specific biology, such as the menstrual cycle, is not covered by existing animal test guidelines. This motivates the development and use of alternative methods that generate human-relevant, predictive information suitable for hazard and risk assessment.

The current status of DART alternatives to animal testing shows comparatively strong coverage of early embryo-fetal development that capture early cell differentiation, gastrulation-like processes, and germ-layer specification and differentiation. Within this domain, two human-based assays are already applied in tiered testing approaches in industrial and pharmaceutical development settings. By contrast, several domains remain under-developed and require targeted research and method development, notably early-gestation placenta biology (first and early second trimester), male and female fertility, and postnatal windows such as lactation, mini-puberty, and puberty. At present, there is no qualified or validated DART *in vitro* battery and no formally agreed structured approach specific to DART. However, stepwise, exposure-led NGRA frameworks have been proposed that organize DART-relevant alternatives to animal testing together with *in silico* tools to support quantitative decision-making (Mueller et al., 2025; Rajagopal et al., 2022). Ongoing efforts to advance human-relevant alternatives to animal testing for DART regulatory testing include dedicated DART initiatives and broader NGRA projects that bring multiple stakeholders across chemical, pharmaceutical and food risk assessment together for consensus building and connect method development with fit-for-purpose use.

On this basis, the chapter outlines short-, medium-, and long-term actions to address scientific gaps and implement measures needed for a paradigm change. The objective is to support and advance the roadmap to phase out animal testing, delivering DART assessment that is more efficient, effective, and humane while maintaining high levels of health protection.

Why a change for DART assessments?

Development and reproduction are major public-health concerns. According to WHO, about 6% of pregnancies worldwide result in congenital birth defects, accounting for ~10% of deaths in children under the age of five (WHO, 2025a). Beyond known contributors, such as genetics, nutrition, infections, and lifestyle, environmental chemicals are increasingly linked to approximately 50% of congenital disorders without a specifically identified cause (WHO, 2025b). At the same time, many countries report declining female and male fertility, and recent projections suggest that by the year 2100, ~97% of countries will have fertility rates below the level needed to sustain their populations (GBD, 2021).

Pharmaceuticals

Because pregnant women are considered a vulnerable population, they are typically excluded from clinical trials. Although limited guidance is available in ICHM3(R2), and additional considerations are provided in the Clinical Trial Coordination Group (CTCG) guideline, further guidance for the inclusion of pregnant women in clinical trials is currently being prepared (ICH E21). As of now, however,

regulatory assessment of DART endpoints for pharmaceuticals still relies primarily on animal studies defined in the ICHS5(R3) guideline. In general, DART is assessed in three separate studies covering the reproductive cycle: the fertility and early embryonic development (FEED) toxicity study, the embryo-fetal developmental (EFD) toxicity study (usually performed in both a rodent and a non-rodent species), and the pre-and postnatal developmental (PPND) toxicity study. In addition, ICHS5(R3) provides qualification criteria for use of DART alternatives to animal testing under defined contexts of use, including a reference compound list of known human teratogenic pharmaceuticals.

Chemicals

For chemical risk assessment, mainly OECD guidance is followed (e.g. OECD TG 414, 421,422 and 443). For a single substance registered under REACH at the highest tonnage band, DART studies can require up to ~4,500 animals (van den Heuvel et al., 2023). In particular, the EOGRTS (TG 443) is highly animal-intensive and, while ECHA rates it as a credible and useful tool for identifying reproductive and developmental hazards, it has also faced substantial criticism over study design choices (e.g. dose selection and cohort triggers), limited endocrine sensitivity, and overall animal use (ECHA, 2023). Moreover, the translational potential of animal studies to humans is limited by substantial species differences. This is especially evident in female reproduction and placenta biology: the human menstrual cycle is not mimicked in standard animal tests, and the placenta is among the most species-specific organs (Armstrong et al., 2017). Furthermore, postnatal periods are only partially represented in current guidelines. In rats, TG 443 covers lactation and selected puberty markers, but mini puberty, which is a transient (postnatal month 1-6) reactivation of the reproductive hormonal system that mirrors puberty on a smaller scale, essential for imprinting sexual and neuroendocrine development, is not specifically assessed. In rabbits, standard guidelines do not include postnatal endpoints (TG 414). Consequently, thousands of chemicals remain inadequately evaluated for developmental and reproductive hazards.

For both the pharmaceutical and chemical fields, adopting alternatives to animal approaches that generate mechanistic, human-relevant evidence for DART while reducing reliance on apical animal endpoints is required. Complementary DART- alternatives to animal testing such as *in silico* methods and *in vitro* approaches for hazard identification, together with toxicokinetic and exposure modelling (PBPK/QIVIVE) need to be integrated into structured, fit-for-purpose testing strategies for risk assessment. Over recent years, international efforts (Health and Environmental Sciences Institute, 2025; Schenk et al., 2010) have accelerated the development, evaluation, and alignment of DART-relevant alternatives to animal testing for regulatory use, laying the groundwork for more consistent implementation across EU frameworks.

DART *in vitro* non-animal approaches

DART spans female and male reproduction, placenta development and function, embryo-fetal development, and the postnatal periods including lactation, and puberty (Figure 17), which are individual endpoints with very different molecular and cellular physiology. To support decision-making, coherent batteries using non-animal approaches are needed to provide sufficient biological coverage across these phases, processes, and crucial pathways to ensure protection, potentially with domain-specific batteries (e.g. male reproduction, female reproduction, and development). Once established, these batteries can serve as one tier within a broader tiered approach (see example DNT). Biological coverage by *in vitro* alternatives to animal testing has advanced markedly for early embryo-fetal development through stem-cell-based platforms, enabled by human iPSC/ESC systems that can differentiate towards multiple embryonic lineages (ectoderm, mesoderm, endoderm) and extra-embryonic lineages (e.g. trophoblast) (Shahbazi & Pasque, 2024). In contrast, other areas remain comparatively under-developed and need targeted method development, notably placenta biology,

female reproduction, and the postnatal phases (lactation, postnatal and child development, puberty), as illustrated in Figure 17.

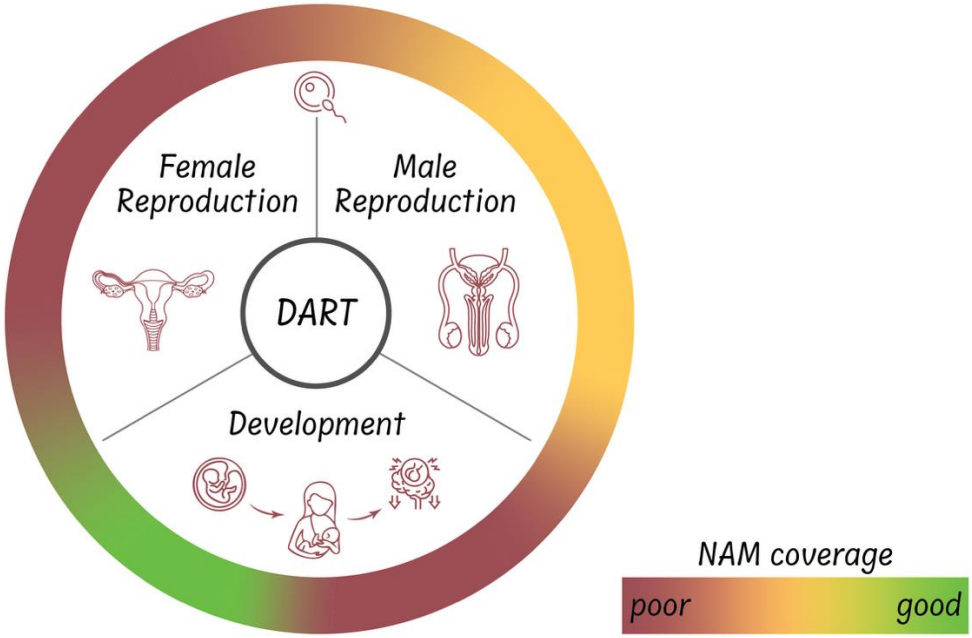


Figure 17: Schematic representation of key developmental processes and corresponding to alternative approaches (or new approach methodology (NAM)) coverage.

Female reproduction

One area requiring further research and development of alternatives to animal testing is female fertility, which is particularly important given substantial inter-species differences in oocyte maturation, menstrual cycling, and menopause (Caligioni, 2009). Notably, neither rodents nor rabbits menstruate or exhibit a human-like menopause. Likewise, the placenta, central to healthy embryonic development, remains under-addressed. Beyond acting as a barrier, the placenta carries essential metabolic, endocrine, and immunological functions, and perturbations can trigger adverse effects on the developing child. Because placenta development, morphology, and function differ markedly across species, not only aspects of female reproduction, but also human-specific key features of placenta biology are not adequately represented in traditional animal tests. Although some *in vitro* models for female reproduction and placental tissues exist (e.g., implantation, menstrual-cycle models, placental organoids, *ex vivo* placental models), they are primarily used for basic research and have not been developed for application in a regulatory context (Garcia-Alonso et al., 2021; Li et al., 2024).

Male reproduction

The most common *in vitro* models for male reproductive systems target the testis and prostate. Within the testis, recent advances focus on Leydig, Sertoli, and germ cell functionality, emphasizing steroidogenesis, spermatogenesis, and testicular development. While the H295R steroidogenesis assay (OECD TG 456) is the only validated *in vitro* method for steroidogenesis testing, its suitability for assessing gonad steroidogenesis is frequently questioned due to its derivation from a cancerous, adrenal cell line of female origin (Botteri Principato et al., 2018). Human-based alternatives to animal testing currently in the refinement and optimization stage include human testis organoids (Madureira Silva et al., 2025) that capture the structural and functional complexity of the testis, as well as hiPSC-derived Leydig-like models aimed at studying steroidogenesis (Sato et al., 2025). More complex alternatives to animal testing under development include early liver–testis multi-organ chips that couple biotransformation with germ-cell toxicity, and ongoing efforts to refine robust differentiation protocols for hiPSC-derived spermatogonia as well as advanced 3D cultures that reconstruct an intact blood–testis barrier for transport studies.

Development

Currently, three DART alternatives to animal testing (Embryonic Stem Cell Test (EST), Whole Embryo Culture (WEC), and the Micromass assay) have been validated by ECVAM. These methods rely on animal-derived cells or *ex vivo* embryo culture. Many additional models are available and in various stages of development, however, a detailed landscaping of these approaches, while clearly needed, is beyond the scope of this document. Importantly, human-based systems offer clear advantages in biological relevance, but their qualification or validation remains challenging due to the limited availability of well-characterized human reference compounds. As a result, human and animal models need to be developed and used in parallel to ensure robust coverage of developmental processes.

Two commercially available but not yet validated assays focus on early developmental phases and processes: ReproTracker addresses germ-layer differentiation (Jamalpoor et al., 2022), and devTOX quickPredict assesses stem-cell proliferation (Palmer et al., 2013). However, they do not cover later stages such as organogenesis and maturation. Ectodermal neurodevelopment is well supported by the DNT *in vitro* battery (DNT IVB; OECD, 2023) and other assays, like the UKN1 (Seidel et al., 2022; Krug et al., 2012), the RoFA assay (Dreser et al., 2020), or the Neurosetta assay (Lundin et al., 2024) studying neural tube formation, are being developed. Mesodermal and endodermal lineage coverage beyond early differentiation remains under-developed, and targeted method development is needed to address organogenesis and functional maturation in these domains. Prioritisation can happen through mapping with human developmental diseases and relevant pathways.

Endocrine domain

Endocrine pathways play a central role in female and male reproduction as well as in pre- and postnatal development, making alternatives to animal testing that capture endocrine activity essential for DART assessment. Compared to other DART domains, endocrine modalities, particularly the estrogen, androgen, thyroid, and steroidogenesis (EATS) pathways, are relatively well served by existing *in silico* and *in vitro* tools (EFSA, 2018), including two validated assays: the H295R steroidogenesis assay (OECD TG 456) and the estrogen receptor binding assay (OECD TG 493). Non-EATS pathways are partially covered within the DNT-IVB (Koch et al., 2025), and important endocrine processes outside the established modalities, such as those along the hypothalamic-pituitary-gonadal (HPG) axis, still lack human-relevant models. Moreover, current approaches often focus on agonist activity, whereas antagonistic mechanisms and more complex modes of endocrine disruption remain challenging to assess. Continued method development is therefore needed to ensure more complete and mechanistically informative coverage of endocrine pathways relevant to DART.

Overall, the biological coverage of toxicity endpoints across existing DART alternatives to animal testing is not yet fully resolved. Defining the context of use, including the chemical applicability domains and biological applicability domains of these test systems is crucial to understand their use in the regulatory setting. Although many genes and pathways relevant to fetal development have been identified (Janowska-Sejda et al., 2022), the development of DART-relevant AOPs remains challenging and substantial effort needs to be spent in this area. Systematic mapping is therefore crucial to (i) delineate coverage of alternatives to animal testing across life stages, (ii) identify context of use and applicability domains of specific alternatives to animal testing and battery approaches, and (iii) establish best-fit combinations for first tier (screening) testing that can flag clear positives and low-concern cases before more refined higher-tier assessments are considered.

Mapping Initiatives toward regulatory uptake of DART alternative to animal testing

This section summarises ongoing initiatives and projects that support the uptake of DART-relevant methods for regulatory decision-making. Building on the increasing landscape of DART-relevant alternatives to animal testing , their emerging readiness, and the gaps identified earlier, several efforts have been set up to support and accelerate regulatory uptake. These efforts include dedicated DART groups and initiatives as well as broader next-generation risk assessment (NGRA) projects that include DART activities, such as the EU projects PARC and CHIASMA. Together, they aim to connect method development with fit-for-purpose use in structured assessment approaches.

The initiatives listed below are among the major ongoing efforts:

HESI DART

The HESI Developmental and Reproductive Toxicology (DART) Committee brings together scientists from industry, government, and academia to exchange information, advance DART science, and build consensus on how experimental evidence supports human health risk assessment (Health and Environmental Sciences Institute, 2025). This group consists of industry, regulatory and academic representatives across chemical, pharmaceutical, cosmetics and food industries. Typical activities include workshops, consensus-building discussions, and publications aimed at improving how DART evidence is generated and used.

Among several topic-focused working groups, the **DART alternatives to animal testing /Alternatives** working group concentrates on creating a Non-Animal Method toolbox and clarifying contexts of use for alternative assays to meet regulatory expectations and progress toward validation and use. The working group organised the hybrid workshop “NAMs in DevTox Testing: Moving the Needle Forward to Regulatory Use” held 30th September – 1st October 2025 (Washington, DC & online), focused on advancing regulatory application of DART alternatives to animal testing .

In addition, the **DARTable Genome** working group integrates chemical–protein target information with pharmacokinetics and developmental biology to link modes of action to DART outcomes and support data-driven IATAs. Ongoing work includes a conceptual framework and case studies with well-characterised teratogens to illustrate how such pathway knowledge can inform PoDs and regulatory interpretation (Janowska-Sejda et al., 2022).

Furthermore, this group aims to reduce animal testing by investigating the use of weight of evidence approaches to waive animal testing, by performing retrospective research, including investigating the need of testing in a second species for small molecules (Theunissen et al., 2016), and ways to reduce NHP (Bowman et al., 2026; Roos et al., 2025) testing for DART.

ISTNET-DART

The International STakeholder NETwork for DART (ISTNET-DART) brings together regulators, international institutions, companies, and universities to advance mechanism-driven, human-relevant alternatives to animal testing for DART regulatory testing. Building on the successful example of ISTNET-DNT, which led to an OECD recommendation for a DNT *in vitro* battery (DNT-IVB), DART assays are intended to be based on the underlying biology of human reproduction and development. The initiative began with an in-person workshop in Zurich (12–13 September 2024) with 61 participants from 12 countries, followed by online workshops in February and April 2025 (Fritsche et al., 2024). The Zurich meeting featured presentations on the biology of the reproductive and developmental cycles and the status of *in vitro* methods for DART testing, including physiology-based kinetic (PBK) modelling, and was followed by breakout groups on female and male reproductive toxicity, developmental toxicity, and PBK modelling. Participants agreed that methods should be placed in the context of AOPs for DART or measured through adverse responses verified within the Non-Animal Method by selected benchmark substances. A meeting report has been published, and a long report including a roadmap is in preparation.

Proposed Frameworks and Case Studies

At present, there is no regulator-endorsed structured approach for DART alternatives to animal testing . However, some publications have proposed ways to organise DART-relevant alternatives to animal testing for decision-making. Approaches vary in maturity: some offer conceptual, tiered structures using AOP-guided integrations and decision trees structures (Burbank et al., 2023; Catlin et al., 2024; Hareng et al., 2005), while others implement and test stepwise approaches quantitatively.

A recent case study demonstrated an adaptable approach for the assessment of chemicals with limited traditional toxicity data, leveraging a Non-Animal Method-based assessment workflow (Paul Friedman et al., 2025). The study specifically incorporated DART assessment by combining *in silico* (COMPARA, CERAPP, TEST DEV) and *in vitro* (ToxCast ER and AR models, devTOX quickPredict) alternatives to animal testing to develop DART hazard flags, which serve as qualitative, putative indicators of hazard related to developmental toxicity and endocrine modulation.

Quantitative, exposure-led assessments for DART utilize stepwise NGRA frameworks that integrate *in silico* tools and human-relevant alternatives to animal testing. These frameworks link Non-Animal Method-derived hazard information (Point-of-Departure, or PoD) with exposure estimates to calculate the Bioactivity:Exposure Ratio (BER), which serves as an explicit decision criterion for risk prioritization. Rajagopal et al. (2022) introduced a DART Non-Animal Method "core toolbox" and evaluated its coverage against Developmental and Reproductive Signalling (DARS) markers related to human development, finding nearly 80% coverage of DARS genes by the mechanistic tools used. Subsequently, Mueller et al. (2025) evaluated such a framework using 37 benchmark compounds across 49 exposure scenarios. This framework employed a Tier 0 *in silico* module (e.g., Derek Nexus, OECD QSAR Toolbox, VEGA, OPERA) and a Tier 1 PoD/BER module. The Tier 1 module combined broad screening tools, notably High-Throughput Transcriptomics (HTTr) and Cell Stress Profiling (CSP), with targeted DART Alternatives to animal testing, including *In Vitro* Pharmacological Profiling (IPP), the hiPSC-based ReproTracker and devTOX quickPredict assays, and endocrine assays (H295R

steroidogenesis and ER/AR CALUX pre-screens). Using a conceptual BER threshold of 1 to flag high/uncertain risk scenarios, the approach successfully identified 17 of the 18 high-risk exposure scenarios confirmed by existing regulatory opinions. Key limitations identified include the need to establish an agreed-upon protective BER threshold and address remaining gaps in biological coverage necessary for comprehensive DART protection.

Together, these studies illustrate how DART-relevant Alternatives to animal testing, exposure estimation (including pregnancy/fetal considerations), and explicit decision criteria (e.g. BER) can be combined and transparently tested with reference scenarios.

Summary Action Plan

The following short-, medium-, and long-term actions set out the requirements for a staged transition towards implementing alternatives to animal testing in DART assessment.

Short-Term Actions

Policy

- Raise political attention to the need to prioritise DART alternatives to animal testing due to the fundamental societal significance
- Acquire targeted funding to foster innovation in DART alternatives to animal testing and establish frameworks with high predictivity (biology, toxicology, pathology) and feasibility (cost, time, efforts)
- Secure funding for case studies that apply available DART alternatives to animal testing in a WoE approach

Science

- Perform a comprehensive landscaping of existing DART-relevant Alternatives to animal testing
- Identify gaps across female/male reproduction, placenta biology, embryo-fetal development (EFD), and pre- and postnatal phases (lactation, mini-puberty, puberty)
- Establish AOPs and KE-AO relationships to support integration of non-animal method evidence
- Assess readiness of existing alternatives to animal testing and describe their context of use (including biological and chemical applicability domain). Identify methods ready for regulatory support, for instance to support WoE assessment for expected teratogens.
- Start qualifying (e.g. ICHS5(R3)) or validating (chemicals) advanced DART Alternatives to animal testing
- Develop and refine QSAR/read-across models for DART-relevant endpoints

Best Practices

- Develop a strategic roadmap towards animal-free DART testing
- Establish exchange between DART initiatives and related projects to ensure synergies and joint development
- Adapt and operationalise DART-specific readiness criteria
- Expand the reference compound list as presented in ICHS5R3 Develop guidance for mechanistic WoE-based approach for DART assessment, based on e.g. lessons learned from (OECD) IATA case studies testing WoE-based approaches.

Communication

- Develop communication roadmap encompassing the needs identified for policy roadmap, science and best practices, differentiated where needed
- Engage more stakeholders and strengthen international stakeholder exchange to align goals
- Organise platforms for discussion and exchange

Medium-Term Actions

Policy

- Adapt regulatory processes and legislations in different sectors to enable integrated Non-Animal Method-based approach for DART assessment
- Secure targeted funding programmes to support development of missing Alternatives to animal testing, further development of existing alternatives to animal testing and translational activities
- Ensure continuous dialogue between submitters and assessors via safe spaces during the assessment process.
- Establishment of an international expert group integrating different stakeholder perspectives on DART.

Science

- Continue establishment, qualification, and validation of DART alternatives to animal testing to close applicability domains and reduce uncertainty to a level where negative results are reliable for regulatory decision-making
- Build an integrative framework that leverages existing Alternatives to animal testing, defining clear contexts of use and applicability domains
- Conduct read-across case studies on the basis of DART *in vitro* non-animal approaches, designed in co-creation with stakeholders, to increase the acceptance of read-across for DART assessment.

Best Practices

- Create a living catalogue of ready-to-use DART alternatives to animal testing with contexts of use and applicability domains
- Organise inter-laboratory validation tests for alternatives to animal testing that are pivotal for IATA
- Define and establish key elements of quality (control) of Alternatives to animal testing
- Perform uncertainty assessment of Non-Animal Method-based approach for DART assessment and compare to traditional approach, and define level of protection
- Use sector-specific cases to optimize data interpretation and methodologies
- Ensure transferability and public availability of alternatives to animal testing through training for all relevant stakeholders for validation and onboarding of new Alternatives to animal testing

Communication

- Ensure continuous communication between stakeholders (industry, regulators, academia, politics)
- Communicate benefits and contexts of use of the initial framework

Long-Term Actions

Policy

- Secure funding for applying best science in regulatory toxicology (translation of basic science methods into use for regulation)
- Implementation of IATA/battery tool that is ready for regulatory use

Science

- Achieve full replacement of DART assessments with alternatives to animal testing
- Ensure that the DART Non-Animal Method approach follows the evolvement of science by including scientific progress on alternatives to animal testing into existing frameworks

Best Practices

- Ensure alignment of global guidance documents within regulatory pillars (e.g., pharmaceuticals or chemicals) to support coherent approach for DART assessment.
- Review progress and learnings of newly implemented animal-free approach for DART assessment.
- Implement training of next generation professionals (including students, regulators, industry) on new approaches for DART assessment.

Communication

- Integrate methods and concepts into the curriculum of schools and universities.

Conclusion

The actions outlined here represent a set of priorities that will be refined as methods, data, and evaluation criteria evolve. Their delivery depends on sustained, coordinated progress and continuous exchange among stakeholders including policy makers, regulators, method developers, and assay users. The unifying goal is to establish an animal-free framework for DART assessment that is usable across sectors and ensures a level of protection that is safe for human development and reproduction, spanning male and female fertility through prenatal and postnatal development to sexual maturation.

References

Armstrong, D. et al. The core transcriptome of mammalian placentas and the divergence of expression with placental shape. Placenta 57 (2017). https://doi.org/10.1016/j.placenta.2017.04.015

Botteri Principato, N. L., Suarez, J. D., Laws, S. C. & Klinefelter, G. R. The use of purified rat Leydig cells complements the H295R screen to detect chemical-induced alterations in testosterone production. Biology of reproduction 98 (2018). https://doi.org/10.1093/biolre/iox177

Bowman, C.J. et al. Minimizing use of non-human primates to inform risk to fertility and of adverse pregnancy outcomes with pharmaceuticals. Regulatory Toxicology and Pharmacology 164 (2026). https://doi.org/10.1016/j.yrtph.2025.105971

Burbank, M. et al. Advancing the use of new approach methodologies for assessing teratogenicity: Building a tiered approach. Reproductive toxicology (Elmsford, N.Y.) 120 (2023). https://doi.org/10.1016/j.reprotox.2023.108454

Caligioni, C. Assessing Reproductive Status/Stages in Mice. Current protocols in neuroscience (07/2009). https://doi.org/10.1002/0471142301.nsa04is48

Catlin, N. et al. New approach methodologies to confirm developmental toxicity of pharmaceuticals based on weight of evidence. Reproductive toxicology (Elmsford, N.Y.) 129 (2024 Oct). https://doi.org/10.1016/j.reprotox.2024.108686

Dreser, N. et al. Development of a neural rosette formation assay (RoFA) to identify neurodevelopmental toxicants and to characterize their transcriptome disturbances. Archives of toxicology 94 (2020). https://doi.org/10.1007/s00204-019-02612-5

ECHA. Evaluating results from 55 extended one-generation reproductive toxicity studies under REACH: Final report of the EOGRTS review project. (2023). https://doi.org/10.2823/92503

EFSA. Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009. EFSA Journal 16 (2018). https://doi.org/10.2903/j.efsa.2018.5311

Fritsche, E., Aspiroz, L., Arand, M., Faustman, E. & Müller, I. International STakeholder NETwork (ISTNET) Workshop for creating a developmental and reproductive toxicity (DART) testing roadmap for regulatory purposes. ALTEX 41 (2024). https://doi.org/10.14573/altex.2410081

Garcia-Alonso, L. et al. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. Nature Genetics 53 (2021). https://doi.org/10.1038/s41588-021-00972-2

GBD 2021 Fertility and Forecasting Collaborators. Global fertility in 204 countries and territories, 1950-2021, with forecasts to 2100: a comprehensive demographic analysis for the Global Burden of Disease Study 2021. Lancet (London, England) 403 (05/18/2024). https://doi.org/10.1016/S0140-6736(24)00550-6

Hareng, L., Pellizzer, C., Bremer, S., Schwarz, M. & Hartung, T. The Integrated Project ReProTect: A novel approach in reproductive toxicity hazard assessment. Reproductive Toxicology 20 (2005/09/01). https://doi.org/10.1016/j.reprotox.2005.04.003

Health and Environmental Sciences Institute. Developmental and Reproductive Toxicology (DART), <https://hesiglobal.org/developmental-and-reproductive-toxicology-dart/> (2025).

Jamalpoor, A. et al. A novel human stem cell-based biomarker assay for in vitro assessment of developmental toxicity. Birth defects research 114 (11/15/2022). https://doi.org/10.1002/bdr2.2001

Janowska-Sejda, E., Adeleye, Y. & Currie, R. Exploration of the DARTable Genome- a Resource Enabling Data-Driven NAMs for Developmental and Reproductive Toxicity Prediction. Frontiers in toxicology 3 (01/19/2022). https://doi.org/10.3389/ftox.2021.806311

Koch, K. et al. Nuclear hormone receptors control fundamental processes of human fetal neurodevelopment: Basis for endocrine disruption assessment. Environment International 198 (2025/04/01). https://doi.org/10.1016/j.envint.2025.109400

Krug, A. K. et al. Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach. Archives of Toxicology 87 (2012). https://doi.org/10.1007/s00204-012-0967-3

Li, Q., Turco, M. & Moffett, A. Trophoblast organoid systems to study human placentation. Developmental cell 59 (03/25/2024). https://doi.org/10.1016/j.devcel.2024.02.009

Lundin, B. et al. RosetteArray® Platform for Quantitative High-Throughput Screening of Human Neurodevelopmental Risk. bioRxiv : the preprint server for biology (06/20/2024). https://doi.org/10.1101/2024.04.01.587605

Madureira Silva, S. et al. A new human in vitro model of cytotypic and testosterone-producing organoids derived from testicular tissue of transgender women. Human reproduction open 2025 (08/20/2025). https://doi.org/10.1093/hropen/hoaf043

Mueller, I. et al. An advancement in developmental and reproductive toxicity (DART) risk assessment: evaluation of a bioactivity and exposure-based NAM toolbox. Frontiers in toxicology 7 (06/30/2025). https://doi.org/10.3389/ftox.2025.1602065

OECD. Initial Recommendations on Evaluation of Data from the Developmental Neurotoxicity (DNT) In-Vitro Testing Battery. (2023).

Palmer, J. et al. Establishment and assessment of a new human embryonic stem cell-based biomarker assay for developmental toxicity screening. Birth defects research 98 (2013). https://doi.org/10.1002/bdrb.21078

Paul Friedman, K. et al. Integration of new approach methods for the assessment of data-poor chemicals. Toxicological Sciences 205 (05/01/2025). https://doi.org/10.1093/toxsci/kfaf019

Rajagopal, R. et al. Beyond AOPs: A Mechanistic Evaluation of NAMs in DART Testing. Frontiers in Toxicology 4 (2022). https://doi.org/10.3389/ftox.2022.838466

Roos, P. et al. Retrospective evaluation of the use of non-human primates to inform risk of male and female fertility with pharmaceuticals. Regulatory Toxicology and Pharmacology (2025). https://doi.org/10.1016/j.yrtph.2025.105990

Rovida, C. & Hartung, T. Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals - a report by the transatlantic think tank for toxicology (t(4)). ALTEX 26, 187–208 (2009).

Sato, K. et al. Efficient differentiation of human iPSCs into Leydig-like cells capable of long-term stable secretion of testosterone. Stem cell reports 20 (02/11/2025). https://doi.org/10.1016/j.stemcr.2024.102392

Schenk, B. et al. The ReProTect Feasibility Study, a novel comprehensive in vitro approach to detect reproductive toxicants. Reproductive toxicology 30 (2010). https://doi.org/10.1016/j.reprotox.2010.05.012

Seidel, F. et al. High Accuracy Classification of Developmental Toxicants by In Vitro Tests of Human Neuroepithelial and Cardiomyoblast Differentiation. Cells 11 (2022). https://doi.org/10.3390/cells11213404

Shahbazi, M. & Pasque, V. Early human development and stem cell-based human embryo models. Cell stem cell 31 (10/03/2024). https://doi.org/10.1016/j.stem.2024.09.002

Theunissen, P.T. et al. Comparison of rat and rabbit embryo-fetal developmental toxicity data for 379 pharmaceuticals: on the nature and severity of developmental effects. Crit Rev Toxicol 46 (2016/11). https://doi.org/10.1080/10408444.2016.1224807

van den Heuvel, C., Klaver, N., Tonk, I., Coder, P. & Beekhuijzen, M. Is there any added value of the second generation in the Extended One-Generation Reproductive Toxicity Study (EOGRTS)? A retrospective analysis of 24 EOGRTS. Reproductive toxicology (Elmsford, N.Y.) 122 (10/27/2023). https://doi.org/10.1016/j.reprotox.2023.108493

WHO (2025a). Congenital disorders - Overview, <https://www.who.int/health-topics/congenital-anomalies#tab=tab_1>

WHO (2025b). Congenital disorders - Causes, <https://www.who.int/health-topics/congenital-anomalies#tab=tab_2>

4.7. Endocrine Disrupting Chemicals - Use of New Approach Methodologies for their assessment within European Union regulatory frameworks: Recommendations from the 2024 EPAA Partners' Forum

Introduction

Objective of the Partners Forum

This chapter describes the main conclusions of the European Partnership for Alternative Approaches to Animal Testing (EPAA) Partners' Forum (PF) held on 14-15 November 2024 in Brussels on the use of non-animal approaches for endocrine disruption assessment.

The importance of NAMs in regulatory Endocrine Disruption (ED) assessment

In line with the regulatory frameworks implemented within the EU to reduce the levels of Endocrine Disrupting Chemicals (EDCs) for consumers, workers, and the environment, it was critical to establish validated tests, allowing regulatory authorities to conduct hazard and risk assessments and make relevant decisions (European Commission, Towards a comprehensive European Union framework on endocrine disruptors, 2018; EFSA Scientific Committee, 2013). Many of these methods are now available for the regulatory assessment of EDCs (Council Regulation (EC) No 440/2008). Regulatory measures were taken in the EU with Member States and the EC restricting the use of putative EDCs.[86]

According to the WHO/International Programme on Chemical Safety (IPCS), an EDC is defined as "an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub) populations" (2002). The identification of endocrine disrupting properties of a substance demands according to the definition to demonstrate adverse effects in an intact organism, requiring the use of animal studies. In order to shift away from this paradigm based on animal testing, further development and application of emerging technologies is needed to reduce time and better protect humans and the environment.

EU regulatory frameworks

The EU regulatory framework evolved over the years with specific provisions for the identification and regulation of ED in place in relevant pieces of legislation. Provisions for this purpose are included in the EU Regulation on Classification, Labelling and Packaging of chemicals,[87] the regulation on chemicals in general ("REACH Regulation"), on plant protection products,[88] biocidal products,[89] medical devices,[90] water[91] and in the revised legislation on toys.[92] The requirements vary depending on

[86] Endocrine disruptors - Public Health - European Commission

[87] European Union. Amendment to Regulation (EC) No 1272/2008 as regards hazard classes and criteria for the classification, labelling and packaging of substances and mixtures. 2022. 2023/707

[88] European Parliament. Commission Regulation (EU) 2018/605 of 19 April 2018 amending Annex II to Regulation (EC) No 1107/2009 by setting out scientific criteria for the determination of endocrine disrupting properties. https://eur-lex.europa.eu/eli/reg/2018/605/oj/eng July 23, 2025

[89] European Parliament. Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council. https://eur-lex.europa.eu/eli/reg_del/2017/2100/oj/eng July 23, 2025

[90] European Parliament. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng July 23, 2025

[91] European Parliament. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. https://eur-lex.europa.eu/eli/dir/2000/60/oj/eng July 23, 2025

[92] European Parliament. Regulation of the European Parliament and of the Council on the safety of toys and repealing Directive 2009/48/EC. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52023PC0462 July 23, 2025

the specific legislation. Other legislation, such as that on food contact materials,[93] cosmetics,[94] or protecting workers at the workplace,[95] while not containing specific provisions for ED, allow regulating such substances on a case-by-case assessment. All pieces of legislation make some use of non-animal approach data for the purpose of assessing EDCs.

EU REACH

Currently, REACH information requirements oblige registrants to generate only limited information specifically on endocrine disruption. Under REACH, EDCs can be restricted if a risk has been demonstrated from their use. EDCs can also be identified as substances of very high concern (SVHC) and be subjected to the authorisation procedure. In both procedures, non-animal approaches data can play a prominent role. With the introduction of the new ED hazard classes in classification, labelling and packaging of substances and mixtures (CLP) in 2023, all substances that have been identified before 11 June 2025[96] as SVHCs under REACH due to endocrine disrupting properties will be directly classified as ED Category 1 under CLP.

CLP Regulation and Commission Delegated Regulation (EU) 2023/707 for CLP

The Classification, Labelling and Packaging of chemicals (CLP regulation), supports the increased integrative use of *in vitro* and *in silico* testing methods. In 2023, two new hazard classes for ED, for both human health and the environment, were added in the CLP regulation, in line with the recommendations in the Chemicals strategy for Sustainability (European Commission 2020). Two hazard categories for ED were established based on the level of evidence: known or presumed endocrine disruptors (Category 1) and suspected endocrine disruptors (Category 2), applicable to both human health and the environment. Classification relies heavily on evidence from human and animal data, but there is also a provision for the use of "non-animal data with equivalent predictive capacity". Such data must demonstrate that the substance meets all the following criteria: (a) endocrine activity; (b) an adverse effect in an intact organism or its offspring or future generations; (c) a biologically plausible link between the endocrine activity and the adverse effect (Figure 18). If the evidence conclusively demonstrates that the adverse effects are not relevant to humans (or at the population or subpopulation level), the substance shall not be considered an endocrine disruptor for human health (or the environment). According to CLP, ED is defined as the alteration of one or more functions of the endocrine system caused by an endocrine disruptor (or EDC). *Endocrine activity* is defined as the interaction with the endocrine system that may result in a response of that system, of target organs or target tissues, and that confers on a substance or the mixture the potential to alter one or more functions of the endocrine system. *Adverse effect* is defined as a change in morphology, physiology, growth, development, reproduction or lifespan of an organism, system, population or subpopulation that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress or an increase in susceptibility to other influences. *Biologically plausible link* is defined as the correlation between an endocrine activity and an adverse effect, based on biological processes, where the correlation is consistent with existing scientific knowledge.

[93] European Commission. Commission Regulation (EU) 2023/915 of 25 April 2023 on maximum levels for certain contaminants in food and repealing Regulation (EC) No 1881/2006. https://eur-lex.europa.eu/eli/reg/2023/915/oj/eng July 23, 2025

[94] European Parliament. Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32009R1223 July 23, 2025

[95] European Council. Council Directive on the introduction of measures to encourage improvements in the safety and health of workers at work. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A01989L0391-20081211 July 23, 2025

[96] The deadline extends to 11 June 2026 for substances for which the SVHC-assessment is ongoing, but for which a decision has not yet been adopted under REACH

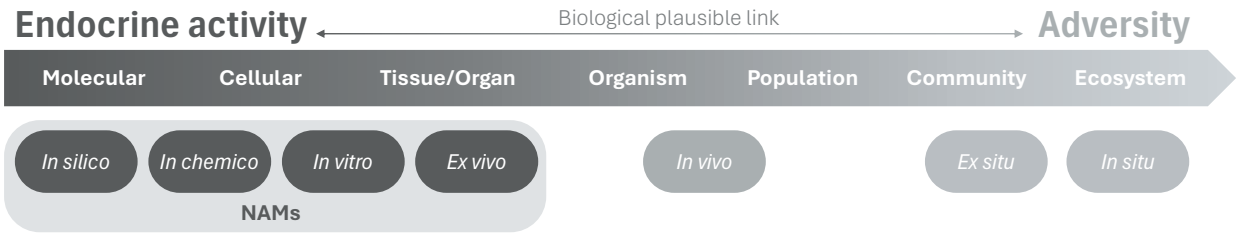


Figure 18. Requirements to classify a chemical substance as endocrine disruptor. Adapted from the Chemical sector perspective presentation by Francesca Trivelloni and the NC3Rs presentation by Natalie Burden in the EPAA PF 2024. NAMs - New Approach Methodologies.

Plant Protection Products and Biocidal Products Regulation.

Commission Delegated Regulation (EU) No 2017/2100 for biocidal products (BPs) and Commission Regulation (EU) No 2018/605 for plant protection products (PPPs) were the first regulations to introduce scientific criteria for the determination of endocrine-disrupting properties for both Human Health and the Environment. To assist the implementation of the criteria, the ECHA/EFSA/JRC 2018 guidance document for the identification of EDs was developed, in the context of Regulations (EU) No 528/2012 and (EC) No 1107/2009, which describes the stepwise process for the scientific assessment of ED properties of BPs and PPPs. The ED assessment includes the integration of regulatory studies as well as public literature studies and different types of evidence including *in vivo*, *in vitro*, *in silico*. More specifically, the ECHA/EFSA/JRC ED Guidance describes how to gather, evaluate and consider all relevant information for the assessment, apply a weight of evidence (WoE) approach and conduct a mode of action (MoA) analysis, in order to help in establishing whether the substance meets (or does not) the ED criteria under the BP and PPP Regulations (ECHA, 2024).

With the introduction of the new ED Hazard Classes in classification, labelling and packaging of substances and mixtures (CLP) in 2023, all substances that have already been identified as EDs under PPPR and BPR will be directly classified as ED Category 1 under CLP. An active substance concluded not to meet the ED criteria under PPPR or BPR, can be classified as Category 1, Category 2 or No classification under CLP, depending on data available when reassessed by the RAC committee.

Organisation for Economic Co-operation *and Development (OECD) Conceptual Framework and Guidance Document* 150

The OECD's work on EDs began in the 1990s, driven by the need of member countries to address the issue of EDCs.[97] To support this effort, the OECD established the Endocrine Disruptor Testing and Assessment Advisory Group (EDTA AG) and other expert groups focused on specific topics and endpoints, such as thyroid function. One significant outcome of this work was the creation of the conceptual framework for testing and assessing EDCs. This conceptual framework organises information across different biological levels, from level 1 (biochemical data) to level 5 (long-term testing on whole organisms) (Figure 19). Levels 1 and 2 do not differentiate between mammalian and non-mammalian tests. Level 4 has a similar number of tests for both mammalian and non-mammalian subjects. Levels 3 and 5 include more non-mammalian than mammalian methods. In total, the conceptual framework includes 44 test guidelines (TGs) and three guidance documents (GDs).

[97] OECD. Endocrine disruptors. https://www.oecd.org/en/topics/sub-issues/testing-of-chemicals/endocrine-disrupters.html (Last accessed: 28 January 2025)

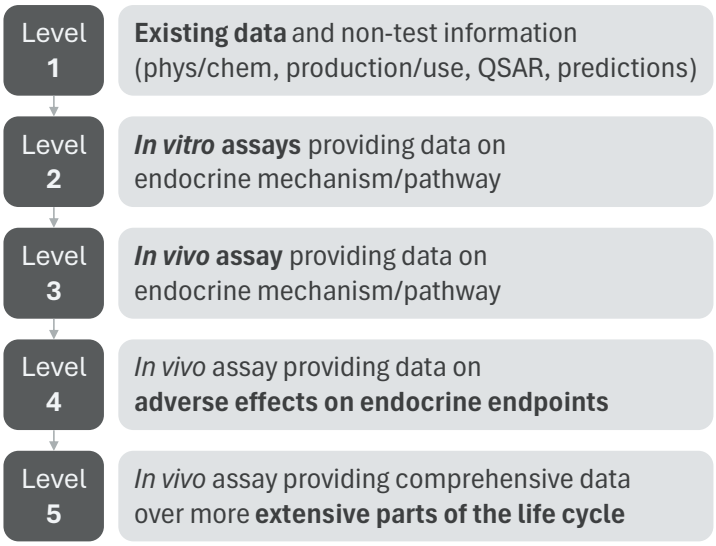


Figure 19. OECD Conceptual Framework. Adapted from the OECD presentation by Elise Grignard in the EPAA PF 2024. QSAR - Quantitative Structure-Activity Relationships. * OECD states (2018): Level 3 - Some assays may also provide some evidence of adverse effects.

The conceptual framework is linked to the Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption, created in 2012 and updated in 2018 (OECD, 2018). This document provides detailed guidance on interpreting test outcomes and compiling evidence to determine if a substance is an EDC. It offers step-by-step instructions for analysing standard test results, weighing evidence for an endocrine activity, and identifying adverse effects in whole organisms. When it was introduced, this guidance was the first comprehensive, international guidance for identifying EDCs. It supports regulatory authorities' decisions related to the hazard of specific chemicals evaluated using test guidelines. The document allows for flexible interpretation to accommodate different domestic legislation, policies, and practices. The document also highlights the importance of not treating the conceptual framework as a linear testing strategy but more as methodology for organising and interpreting data, such that the relative value of studies for different aspects of the assessment can be evaluated.

Research challenges for advancing ED assessments while reducing the reliance on animal testing

The main research challenges mentioned during the EPAA PF were the following:

Endocrine activity vs adversity. Most currently used *in vitro* alternatives to animal testing address endocrine activity (i.e., mode of action) (Figure 18), corresponding to Level 2 of the OECD Conceptual Framework (Figure 19), but they are not yet sufficient on their own to identify endocrine adversity (or lack thereof) at Levels 4-5. Currently, there are no available NAMs that can conclusively determine adverse responses (or lack thereof) at the organism or population level following systemic (acute or repeated) exposure. Developing alternatives to animal testing to demonstrate adversity will take a long time.

Therefore, **short-term recommendations** would be to

(i) Develop approaches that "predict adversity" (or lack thereof) rather than "demonstrate adversity". In this context, it is important to define what "adequately predictive" means.

(ii) Explore combinations of alternatives to animal testing to help predict adversity, particularly in combination with PBK modelling and/or qIVIVE, as single alternatives to animal testing do not provide suitable data to be able to predict adversity. For example, some *in vitro* tests, such as cellular and organoid systems, may have the potential to characterise some adversity (e.g., nervous system cells not developing).

(iii) Explore the potential of the extensive data generated by the pharmaceutical industry on endocrine active pharmaceutical ingredients (APIs) and other compounds with potential side effects. This could be instrumental in creating and completing existing AOPs, through the identification of biomarkers for assessing adverse effects and causation for example.

Uncertainties around negative and positive results. A substance may no longer be considered an EDC under a regulatory framework as if it shows negative results in assays for levels 2 and 3 of the OECD conceptual framework. However, this requires the regulators to have confidence that these results are truly negative. Such a conclusion, however, is only valid for the mechanistic information and mode of actions covered by these assays.

On the other hand, if sufficient data on adversity are available, with the body of evidence indicating an ED-related concern, then the outcome could be "ED criteria are met" even if mechanistic studies are negative. It should also be noted that studies on adversity have limitations with regards to the endpoints covered, which may mean that not every EDC will be identified. It might be considered that the available assays are not sufficient to detect all possible ED-related modes of action, and that uncertainties, such as inadequate concentrations or doses, or unaccounted metabolic factors, can affect the interpretation of results.

Considerations on certainty are similarly relevant for positive results observed in level 2 and 3 assays if such results are used for triggering further *in vivo* testing. For example, a positive *in vitro* assay may not manifest *in vivo* due to metabolism of the molecule to a less (or non-) active form, the dose might not be achievable *in vivo*, or the substance might not reach the target *in vivo*. Depending on the number of level 2 and 3 assays applied in parallel, potentially a high number of unnecessary follow-up testing might be triggered.

Challenges to address:

(i) Building trust in alternatives to animal testing by finding consensus on acceptable uncertainty and predictive capacity to prevent assuming inherent superiority of *in vivo* methods over alternatives to animal testing.
(ii) Develop a solid understanding of the biological mechanisms that allows *inter alia* to determine to what extent mechanisms are covered and may promote widespread use of alternatives to animal testing for the identification of EDs with the ultimate goal to minimise the need of animal testing.
(iii) Build consensus on uncertainties of negative and positive results for both *in vivo* and *in vitro* data.

In conclusion, to prevent assuming inherent superiority of *in vivo* methods over alternatives to animal testing, it is crucial to build trust in the alternatives by finding consensus on acceptable uncertainty and predictive capacity. Widespread use of alternatives to animal testing for the identification of EDCs with the goal to minimise the need of animal testing requires a solid understanding of the biological mechanisms that allows inter alia to determine to what extent mechanisms are covered, but also a thoughtful consideration of the uncertainties of negative and positive results for both *in vivo* and *in vitro* data.

Reduction vs replacement. While some *in vitro* methods can replace animal tests for relatively simple toxicities (e.g., the ToxCast ER pathway model to replace the Uterotrophic assay), in the short term, it may be challenging to replace all *in vivo* tests with alternatives to animal testing, due to the complexity of the whole organism and testing with substances not suitable for culture conditions. To start the transition to alternatives to animal testing, the aim should be to reduce animal testing whilst ensuring equal protection levels. If a test (or battery of tests) is identified that can truly replace an *in vivo* test, efforts should be made to implement it.

Knowledge gaps in alternative methods development

There is a critical need for more knowledge on the impact of EDs at crucial life stages, inter-species differences, and the effects of mixtures. Additionally, only the alternatives to animal testing relevant to

EATS modalities which are validated are currently widely used, despite the existence of other EATS and non-EATS alternatives to animal testing that can address certain elements of other pathways (e.g. EURION Methods Table[98]).

Challenges to address:

(i) Increase knowledge on the impact of EDs at crucial life stages, inter-species differences, and the effects of mixtures.
(ii) Address comprehensively non-EATS modalities which can have human and ecologically relevant impacts.
(iii) Map the available alternatives to animal testing defining applicability domain and their potential regulatory context of use to identify remaining gaps (e.g., through designing specific case studies), defining a testing strategy (e.g., using Integrated Approaches to Testing and Assessment (IATA)), and prioritising the development of those relevant to human health and the environment are essential steps.
(iv) Ensure close collaboration between regulators, researchers and assay developers from the beginning of this process to ensure that alternatives to animal testing under development are scientifically robust and meet the current regulatory needs.

Challenges to address:

(i) Need for predictive quantitative and toxicokinetic *in silico* approaches, such as those using AOPs and PBK models, to extrapolate *in vitro* results to species of interest-relevant external doses used to set safety levels, to characterise the biological plausible link between the mode of action and the observe adverse effect to ultimately communicate hazard and assess risk.
(ii) Implement test standardisation for *in vitro* and *in silico* methods to ensure relevance and reliability of results for hazard and risk assessments and build trust in existing and new alternatives to animal testing.
(iii) Determine non-animal approaches effect concentrations relevant to humans or the environment (i.e., molecular initiating events can be established as effect threshold).
(iv) Investigate multi-organ, multi-endpoint, and multi-species issues in the extrapolation process; not unique to non-animal approaches but relates to many current *in vivo* approaches.
(v) Address the heterogeneity of environmental species sensitivity and the sensitivity of alternatives to animal testing for individual organisms and population effects.

Validation, standardisation or qualification

Challenges to address:

(i) Clearly identify, prioritise and communicate regulatory needs for the development of test methods.
(ii) Facilitate the transition between test development and validation (or standardisation or qualification) to accelerate potential uptake (e.g., PEPPER platform).

Transition to holistic and mechanistic approaches.

Challenges to address:

(i) Building stronger links with the pharmaceutical industry and clinicians by defining mechanistic-based panels of *in silico* and *in vitro* assays to better understand disease mechanisms and facilitate the transition.

[98] EURION Methods table

(ii) Develop methods that provide information on multiple endpoints within the same test to allow multiple regulatory applications.
(iii) Include biomarkers that can help predict or inform on adversity.
(iv) Building a toolbox with holistic and mechanistic-driven assessments to focus resources and testing.

Opportunities for cross-sector collaboration

The main opportunities for cross-sector collaboration that were identified by EPAA PF forum participants to advance non-animal alternatives for ED-assessment were the following:

Make use of internal exposure information or ADME/TK

- Refine the stepwise approach described in the ECHA-EFSA guidance for ED-assessment by including information on internal exposure, potency and life-stage applicability at earlier stages to prevent *in vivo* testing from occurring too early in the decision-making process.
- Proactively include all non-animal approach-based information and exposure information in regulatory decision-making, keeping in mind that hazard identification is required under CLP prior to any risk assessment/management.
- Utilise exposure information more effectively, considering use-dependent data.
- Explore the application of non-animal approach-based potency data for the effective classification and assessment of EDCs. It should be clarified that "potency" in *in vitro* studies is simply an indication of whether the substance is active or not, in order to anticipate activity *in vivo*. Potency is unrelated to the threshold of the adverse effect i.e. a dose over which the adverse effect is observed. This includes exploring the role of potency data in regulatory decision-making, particularly in evaluating activity and redefining adversity by incorporating pathway-specific information.
- Address the challenge of using potency information across species, crucial for both human-relevant and ecological Alternatives to animal testing.

Prioritise non-animal approaches for validation

- Identify promising non-animal approaches that could be validated to address regulatory gaps (e.g. PARC opportunity, linked to EPAA via PARC SYNet, EURION non-EATS Alternatives to animal testing).
- Facilitate the interface between researchers and the regulatory community through the AOP framework and assess the effectiveness of this mechanism.
- Develop case studies/IATA to create non-animal approach batteries, defined approaches, and testing strategies.
- Refine the current stepwise approach to ED assessment to increase flexibility and allow for "exit" steps where appropriate.

Set up regulatory context for using alternatives to animal testing

- Describe the regulatory context of use for available non-animal approaches and map these against different EU regulatory frameworks to identify remaining gaps.
- Build confidence in non-animal approach information demonstrating no ED activity by identifying regulatory examples where non-animal approaches have been used for ED assessment to support the conclusion that ED criteria are not met (e.g., ToxCast ER pathway model) Examples of such confidence building in EFSA's realm are the following: (i) experience from pesticide peer review[99], (ii) development of an adverse outcome pathway network for

[99] Outcome of the pesticides peer review meeting on general recurring issues in mammalian toxicology and ecotoxicology related to endocrine disruptors

reproductive toxicity endpoints to support identification of endocrine disrupters[100], (iii) EFSA database of the outcomes of ED assessments carried out so far[101].https://www.efsa.europa.eu/sites/default/files/documents/ed-report-overview-endocrine-disrupting-assessment-pesticide-active-substances.xlsx

- Ensure that complex non-animal approaches (like e.g. organs-on-chip) are standardised, relevant and reliable and can be used in regulatory decision-making and are not ignored due to their complexity.
- Some participants of the EPAA PF recommended placing greater emphasis on external exposure data when assessing whether a substance may pose a risk for specific uses. This evaluation should ideally occur before deciding to initiate animal testing to determine if the substance is an ED. Such an approach could also reduce the number of false positives from a non-animal approach battery. No consensus was reached on this point. There was a disagreement at the workshop on whether the EU approach for risk management of EDs should be based on hazard identification only or on information on exposure and risk as well.

Collaboration across sectors

- Facilitate ongoing dialogue between researchers, regulators, and industry in a "safe space" (trusted environment) to address scientific gaps and research challenges and build confidence without compromising the regulatory approval process. The creation of expert groups, such as the already established ECHA ED expert groups, could facilitate this dialogue.
- Create a common platform for data sharing, e.g. starting from EFSA's ED database[102] and especially regarding pharmaceutical data, to better demonstrate the relevance of non-animal approaches, particularly in defining biologically meaningful responses and addressing multi-organ and multi-endpoint challenges and predicting adversity. This platform can also help inform on relevant AOPs, review testing strategies, revise guidance, analyse negative non-animal approach results, and extrapolate results across species.
- Set up retrospective case studies across industries, agencies, and researchers to better understand negative non-animal approach results and built confidence when "ED criteria were not met". Use the generated data and insights to inform a revision of the 2018 ECHA/EFSA/JRC ED guidance.
- Create an ED NAM User Forum organised by EPAA to facilitate consensus-building and support increased regulatory use of alternatives to animal testing for ED assessment using real-life dossiers and case studies.

Summary and conclusions

The 2024 EPAA PF highlighted the increasing need in the use of alternatives to animal testing to assess EDs across various industrial sectors. While non-animal approaches can effectively identify endocrine activity, they are not yet sufficient on their own to identify chemicals as EDs since they do not provide the necessary information for assessing adversity. Standardised predictive models and integrating information on internal exposure, potency and life-stage applicability with non-animal approach findings could enhance identification of EDs and risk assessment by improving the link between endocrine activity and adversity. Building trust in both existing and new *in vitro* and *in silico* methods is crucial for the effective implementation of non-animal approaches for EDs. Achieving this will require consensus on the acceptable level of uncertainty for evaluating endocrine activity and adverse effects as well as ensuring methods have predictive capacities equivalent to (or superior to) human, environment, and animal data.Closer dialogue among all actors involved, e.g., regulatory agencies,

[100] https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2026.EN-9974
[101] https://www.efsa.europa.eu/sites/default/files/documents/ed-report-overview-endocrine-disrupting-assessment-pesticide-active-substances.xls
[102] Publicly available EFSA ED DB

academia and industry, is needed to integrate non-animal approaches into safety assessments across regulatory landscapes.

References

EFSA Scientific Committee. Scientific Opinion on the hazard assessment of endocrine disruptors: Scientific criteria for identification of endocrine disruptors and appropriateness of existing test methods for assessing effects mediated by these substances on human health and the environment. EFSA Journal. 2013;11:3132.

European Chemicals Agency. Guidance on the application of the CLP Criteria – Guidance to Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures. Part 3, Health hazards: European Chemicals Agency; 2024.

European Commission. Commission Regulation (EU) 2023/915 of 25 April 2023 on maximum levels for certain contaminants in food and repealing Regulation (EC) No 1881/2006. https://eur-lex.europa.eu/eli/reg/2023/915/oj/eng July 23, 2025

European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Towards a comprehensive European Union framework on endocrine disruptors (2018). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52018DC0734 Accessed July 23, 2025

European Commission. Endocrine disruptors. https://health.ec.europa.eu/endocrine-disruptors_en (Last accessed: 28 January 2025)

European Council. Council Directive on the introduction of measures to encourage improvements in the safety and health of workers at work. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A01989L0391-20081211 July 23, 2025

European Council. Council Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). https://eur-lex.europa.eu/eli/reg/2008/440/oj/eng July 23, 2025

European Parliament. Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council. https://eur-lex.europa.eu/eli/reg_del/2017/2100/oj/eng July 23, 2025

European Parliament. Commission Regulation (EU) 2018/605 of 19 April 2018 amending Annex II to Regulation (EC) No 1107/2009 by setting out scientific criteria for the determination of endocrine disrupting properties. https://eur-lex.europa.eu/eli/reg/2018/605/oj/eng July 23, 2025

European Parliament. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. https://eur-lex.europa.eu/eli/dir/2000/60/oj/eng July 23, 2025

European Parliament. Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32009R1223 July 23, 2025

European Parliament. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng July 23, 2025

European Parliament. Regulation of the European Parliament and of the Council on the safety of toys and repealing Directive 2009/48/EC. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52023PC0462 July 23, 2025

European Union. Amendment to Regulation (EC) No 1272/2008 as regards hazard classes and criteria for the classification, labelling and packaging of substances and mixtures. 2022. 2023/707

European Union. Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). 2006. REGULATION (EC) No 1907/2006

Krewski D, Acosta D, Jr., Andersen M, Anderson H, Bailar JC, 3rd, Boekelheide K, et al. Toxicity testing in the 21st century: a vision and a strategy. J Toxicol Environ Health B Crit Rev. 2010;13:51-138.

OECD. 2018. Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption.

OECD. Endocrine disruptors. https://www.oecd.org/en/topics/sub-issues/testing-of-chemicals/endocrine-disrupters.html (Last accessed: 28 January 2025)

World Health Organization. 2002. Global assessment of the state of the science of endocrine disruptors. Damstra TB, Sue; Bergman, Aake; Kavlock, Robert; Van Der Kraak, Glen (editors).

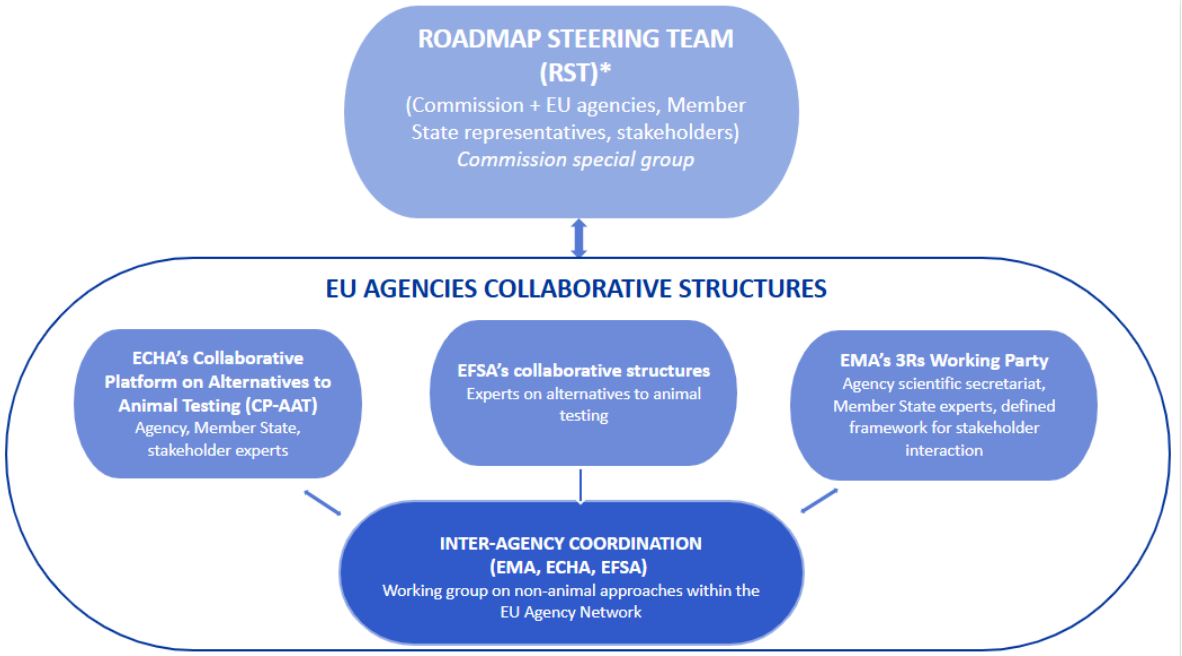
5. Working together to design and implement non-animal approaches in regulatory testing

5.1. Role and composition of the roadmap steering team

Overall description

The core governance structures ensuring implementation of the roadmap will be a Roadmap Steering Team, based on an existing Commission inter-service group (ISG). The Roadmap Steering Team will oversee the implementation of the roadmap and ensure coordination between the other organisational structures supporting the roadmap implementation. Collaborative structures at each of the three EU agencies, ECHA, EFSA and EMA, will play an important role for implementing the roadmap and offer opportunities for stakeholders to get involved. An inter-agency working group within the EU Agency Network (EUAN), consisting only of experts from the three agencies, will allow an efficient coordination between the EU agencies.

These structures are expected to interlink as indicated in Figure 20 below:



*RST action teams will be set up if needed with the option of stakeholder participation

Figure 20: how the main organisational structures will interlink for the roadmap's implementation

5.2. Roadmap steering team

A roadmap steering team will be set up, which will have the role to oversee and steer the actions set out in the roadmap in order to implement the roadmap successfully, provide feedback and suggestions, and activate the members' own networks to ensure that they are widely and actively involved in the roadmap's implementation.

Tasks of the roadmap steering team

- Driving forward implementation of the roadmap;
- Maintaining an overview of the progress of the roadmap's actions;
- Facilitating exchange of views and communication between the different organisational structures implementing the roadmap;
- Initiating and organising high-level communication activities to inform and engage stakeholders, in close collaboration with agencies or other established organisational structures (e.g. the EPAA, the European Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM), etc.);
- Engaging with non-EU partner countries and multilateral organisations to promote phasing out animal testing for chemical safety assessments;
- Providing advice in line with the roadmap's priorities, identifying initiatives that could feed into the roadmap's actions and promoting the catalogue of transitional initiatives.

Composition and working methods

The roadmap steering team will be set up as a Commission 'special group [103]'. It will be composed of the Commission inter-service group members and represent all relevant Commission services, the three EU agencies (ECHA, the EMA and EFSA) and representatives from Member States and different stakeholder organisations. To allow for fruitful discussions, the number of seats for Member States and stakeholders will be limited to a maximum of 12. The following composition is proposed:

- two industry representatives;
- two NGO representatives;
- two representatives of the scientific community;
- three representatives from Member States representing different sectors ;
- further members will be determined by the ISG based on their expertise or function in other organisational structures.

The representatives of the Member States and of the stakeholders will be selected by the ISG in response to a call for expressions of interest. Experience of animal-free approaches and connection with a wider community of stakeholders will be important selection criteria. Members will serve on the roadmap steering team for two years, with the possibility to renew their participation. The team's mandate and working procedures will be determined by the Commission at the beginning of the roadmap's implementation phase. The roadmap steering team is expected to meet regularly (e.g. three to four times a year) to maintain the momentum of the roadmap's implementation.

Immediately after the publication of the roadmap, the Commission will draft a mandate and working procedures for the roadmap steering team and launch a call for participation. After the selection process, the Commission will set up the team and call for a first meeting.

If needed, the Commission will set up action teams to assist the Roadmap Steering Team in advancing particular action points. Stakeholders can be invited to an action team and will be selected according to the Commission's rules [104].

[103] Special groups are bodies that for various reasons do not fit into any existing framework, such as that on expert groups or comitology committees. Special groups do not qualify as 'Commission expert groups' or 'other similar entities' in the sense of Commission Decision C(2016)3301 establishing horizontal rules on expert groups.

[104] Commission Decision C(2016) 3301 establishing horizontal rules on the creation and operation of Commission expert groupshttps://eceuropaeu.sharepoint.com/sites/expert-groups/Shared Documents/Forms/AllItems.aspx?id=/sites/expert-groups/Shared Documents/guidelines-on-the-implementation-of-the-horizontal-rules.pdf&parent=/sites/expert-groups/Shared Documents

5.3. Engaging Member-State and stakeholder experts in agency collaborative structures

EU agency collaborative platforms and informal expert groups - context

The European agencies (ECHA, EFSA and the EMA) have already been working for several years on identifying alternatives to animal methods which could be applied in the regulatory processes they manage. Annex II describes in more detail the work of these three agencies in this area. Nevertheless, it may be necessary to adapt some of their internal organisational structures or working methods to improve and accelerate the implementation of animal-free methods, e.g. setting up or strengthening expert bodies in this specific field.

Stakeholders were consulted on the need to have expert groups dedicated to animal-free approaches within the agencies. A large majority of stakeholders consulted share the view that scientific expert groups [105] are necessary to provide advice on the development of alternative or animal-free approaches and on their uptake and use in a regulatory context [106]. However, there are different views on their role and how these groups should be set up and organised [107].

Some agencies such as the EMA and EFSA have already set up such groups on fostering the adoption of animal-free methods in some of the regulatory fields in their remits. However, these structures do not necessarily cover certain tasks which would be desirable, such as providing non-binding advice on regulatory case studies and guidance for industry to implement alternatives to animal testing [108]. Therefore, these structures might need to be adapted to cover the envisaged tasks described below.

In addition, other agencies like ECHA did not have these types of structures in place dedicated to animal-free testing. During the roadmap development process, ECHA announced the creation of such a new structure – the 'Collaborative Platform on Alternatives to Animal Testing (CP-AAT) – which would gather (i) experts from ECHA staff on alternatives to animal testing and the relevant legislation; (ii) experts from Member-State competent authorities on REACH, CLP and the BPR; and (iii) observers from the Commission and NGOs.

Composition and tasks

The collaborative organisational structures or expert groups will consist of experts from the authorities of the Member States and stakeholders. They will identify regulatory needs for animal-free test methods and contribute, where appropriate, to:

- exchanging views and recommendations with the roadmap steering team on prioritisation of method development, validation, qualification and standardisation activities;
- providing input on the roadmap's implementation to the roadmap steering team;
- serving as a platform for exchange between the Member States, the agency, the Commission and stakeholders on regulatory acceptance of non-animal testing results within the remit of that agency, and as a forum for stakeholders from research, industry or NGOs to give input, while providing the opportunity for stakeholders such as the EPAA to present the outcome of their projects;

[105] The term 'scientific expert group' is used here in an unspecific meaning and should not be confused with a Commission expert group or a committee operated by an agency. Rather, it denotes a group or a collaborative platform with participation of experts in the field.

[106] See the second stakeholder survey report, Section 3.3, available at https://webgate.ec.europa.eu/circabc-ewpp/d/d/workspace/SpacesStore/53fef874-5a58-45fa-b2ac-5e1450fecd9b/download.

[107] See the second stakeholder survey report, Section 3.3, available at https://webgate.ec.europa.eu/circabc-ewpp/d/d/workspace/SpacesStore/53fef874-5a58-45fa-b2ac-5e1450fecd9b/download; the report on stakeholder interviews Section 3.2.1 available at https://webgate.ec.europa.eu/circabc-ewpp/d/d/workspace/SpacesStore/67861f4d-a282-493a-af96-ec47557db0b1/download and the report on the need and feasibility of an advisory scientific committee on non-animal methods available at https://webgate.ec.europa.eu/circabc-ewpp/d/d/workspace/SpacesStore/6448df32-3864-4e0c-9fa7-1b7e910b40e9/download.

[108] See descriptions of existing structures, gap analysis and conclusions in the following report: https://webgate.ec.europa.eu/circabc-ewpp/d/d/workspace/SpacesStore/6448df32-3864-4e0c-9fa7-1b7e910b40e9/download.

- providing non-binding scientific advice on regulatory processes within the remit of the agency, e.g. for ECHA, as regards REACH in relation to (i) non-animal testing in compliance checks, (ii) testing proposals and (iii) substance evaluation or use of non-animal data for CLP-purposes.

Implementation

The Commission will ensure that the decentralised EU agencies and MS can effectively support implementation of the roadmap through the new collaborative structures or that already existing structures are fit for purpose to boost the uptake of alternatives to animal testing in their regulatory work. Since each piece of Union legislation and each EU agency has its own specificities, the EU agencies (ECHA, EFSA, EMA) should assess their current or new organisational structures taking into account their mandate, working methods and resources, and implement measures, if needed, to strengthen or adapt their existing structures. The composition of the current or proposed new groups dedicated to alternatives to animal testing, their tasks, working methods, and necessary resources should be described by the agencies and should serve as the basis for potential adaptations to current practices.

In addition to fostering the uptake of alternatives to animal testing in their own regulatory fields, a key aspect will be for the agencies to put the mechanisms in place within these structures to ensure the cross-fertilisation of best practice and knowledge with the other agencies via the EU Agency Network and with the roadmap steering team. The agencies are encouraged to ensure good communication and efficient exchange of information within and between the agencies.

5.4. EU Agency Network (EUAN) – working group on animal-free approaches

The EU Agencies Network (EUAN) - context

The EU Agencies Network (EUAN) supports the work of the 52 EU decentralised agencies, joint undertakings (JU) and bodies. The Network provides its members with a platform for exchange and cooperation on areas of common interest, enabling them to share their expertise at European level. It also serves as a single point of contact for the EU institutions when addressing cross-cutting issues of general interest to all agencies and JUs. The EUAN comprises sub-networks, working groups and task forces – fostering technical cooperation on specialised topics. They implement the EUAN strategy and annual work programmes and enhance inter-agency collaboration and communication [109].

It is proposed to set up a EUAN working group on alternatives to animal testing to formalise and make visible the already ongoing inter-agency cooperation on alternative approaches and to provide a means for agencies to speak with 'one voice' on phasing out animal testing. A EUAN working group, consisting of experts from EFSA, ECHA and EMA, will facilitate the exchange of information on the development and implementation of alternatives to animal testing across the different pieces of legislation and bodies concerned, helping the EU agencies to keep abreast of the latest developments in regulatory science, and accelerate uptake across the different pieces of legislation.

Proposed tasks of the EUAN working group on animal-free approaches

- Ensure that the EU agencies exchange best practice to foster the use of animal-free methods in their regulatory fields;
- Set priorities and develop actions addressing replacement (or reduction or refinement) in coordination with agency expert groups and the roadmap steering team;
- Provide a means to coordinate the participation of the EU agencies within EU and international fora.

Implementation

[109] See description of EUAN at https://agencies-network.europa.eu/about-network/our-role-governance-and-strategy_en.

The Commission will ask the EU agencies to set up an inter-agency working group on animal-free approaches within the EUAN. The new EUAN working group on animal-free approaches should meet on a regular basis. The input from the agencies' dedicated structures on animal-free methods (e.g. expert groups) will be key for the work of the EUAN working group as hands-on regulatory knowledge will be shared for the benefit of all agencies and regulatory fields.

A representative of the EUAN working group on animal-free approaches might become a member of the roadmap steering team to inform it about the developments within the agencies and participate in other relevant tasks for the overall implementation of the roadmap.

5.5. Translating innovative methods into regulatory applications – validation, standardisation, qualification

Introduction

Studies used for the regulatory assessment of chemicals in the EU often follow the test guidelines (TGs) of the Organisation for Economic Co-operation and Development (OECD). ICH guidelines and and VICH guidelines apply for human medicines and veterinary medicines, respectively. These guidelines support harmonised regulatory testing of safety, quality and efficacy in their respective domains. Other options for harmonising the acceptance of methods include the standardisation process of the International Organization for Standardization (ISO) or that of the European Committee for Standardisation (CEN), which cooperates with the ISO.). Standards are highly relevant for medical devices, for example.

Studies for the regulatory assessments of chemicals are usually carried out by CROs on behalf of industry registrants/applicants for the purpose of satisfying standard information requirements set out in Union legislation and related (agency) guidance/guidelines.

Validation in the context of OECD TG development

Aim of validation

Before a new test method can be accepted (e.g. for use in a regulatory context where the test method's results are used to make regulatory-relevant decisions), the test method must be validated. Validation is a broadly accepted principle, which applies not only to regulatory testing but to all areas of experimental science. Typically, validation consists of demonstrating that the method's results are relevant for a specific purpose (i.e. the results provide us with information that is relevant for taking the decision) and reliable (i.e. there is consistency in the method's results, when the same chemical is tested over time and/or in different places / laboratories).

This SWD is focused on the assessment of the safety of chemicals in a regulatory context. In the Communication and in this accompanying SWD, the term 'validation study' will be used for short for 'study to validate a method for the assessment of the safety of chemicals'.

Over the past 25 years or so, most of the regulatory-relevant validation studies have been conducted with the aim of developing OECD TGs. As a result, 'validation of method' is – in the context of chemical safety testing – often wrongly understood as 'development of an OECD TG', even though this conflates the scientific process of validation with the more political one of acceptance among OECD members. Indeed, OECD Guidance Document 34 on method validation[110] explicitly recognises these two parts of 'OECD' validation: 1) the scientific establishment of the relevance and reliability of the method's results, and 2) the acceptance of the method as an OECD TG (by a consensus-searching process among the expert representatives of the regulatory authorities of the OECD member countries). This obviously is one of the great advantages of the validation process according to OECD GD34 and of the resulting TGs. Likewise, 'formal validation' is often understood as 'validation according to OECD Guidance Document 34'. Consequently, a formal validation study would not be undertaken unless there were sufficient international regulatory interest to develop a TG based on the method to be validated. The degree of international regulatory interest is assessed through the annual project submission and commenting process managed by the OECD working party of national coordinators for the test guidelines programme (WNT). Prior to this assessment, within the EU, initial proposals for method validation and regulatory application can be submitted for a preliminary opinion on regulatory relevance to the Preliminary Assessment of Regulatory Relevance network (PARERE)[111] – a network of cross-

[110] https://www.oecd.org/content/dam/oecd/en/publications/reports/2005/08/guidance-document-on-the-validation-and-international-acceptance-of-new-or-updated-test-methods-for-hazard-assessment_2695fa12/e1f1244b-en.pdf

[111] https://joint-research-centre.ec.europa.eu/projects-and-activities/reference-and-measurement/european-union-reference-laboratories/eu-reference-laboratory-alternatives-animal-testing-eurl-ecvam/alternative-methods-toxicity-testing/advisory-and-consultation-bodies/parere-eurl-ecvam-network_en

sector EU regulators coordinated by the JRC – via the EURL ECVAM test submission process. In addition, the International Cooperation on Alternative Testing Methods (ICATM) often discusses promising methods and evaluates their relevance, with a view to cooperating on validation and supporting the OECD process.

The system of validation in the context of OECD TG development has significant advantages for both regulators and industry, including: (i) the linking of standard information requirements with standard accepted tests; (ii) the quality assurance of data; (iii) business predictability; (iv) some degree of legal certainty; (v) efficiency in submission and assessment processes; and (vi) reuse of data between sectors and in different international jurisdictions. This reuse of data is possible due to the OECD agreement on mutual acceptance of data (MAD), and applies to data generated using TGs under the strict quality assurance conditions of good laboratory practice [112] (GLP). The development of an OECD TG requires consensus within the OECD WNT that there is enough regulatory need for the new TG across OECD member countries, as well as sufficient confidence in the proposed method. The system also has some disadvantages, including (i) the time taken to evaluate and adopt new methods; (ii) the need to gain consensus (unanimity) within the WNT; and (iii) the lack of accountability and transparency in cases where member countries delay or block the adoption process.

Conducting validation

Since the late 1990s, most formal validation studies for methods assessing the safety-relevant properties of chemicals conducted in the EU have been coordinated by the EURL ECVAM. Studies usually involve close cooperation among method developers (e.g. companies, academic labs, research institutes), industry end users (e.g. large companies, company associations), and the EU network of over 30 validation labs proposed by EU member countries (EU NETVAL), coordinated by the EURL ECVAM. Coordination of a validation study includes its overall design, establishing and facilitating a validation management team, ensuring sufficient resources, reviewing and approving study plans and test protocols, providing day-to-day advice and troubleshooting, gathering and analysing data, producing a validation study report, and organising a peer review by ECVAM's Scientific Advisory Committee [113]. The outcome of the independent peer review marks the end of the formal validation process. The status of method validation and acceptance can be followed using the tracking system for alternative methods towards regulatory acceptance (TSAR). Usually, the EURL ECVAM – on behalf of the Commission – also supports subsequent TG development at the OECD.

Over the past 20 years, EU framework programmes have invested close to EUR 1.5 billion directly into the development of alternatives to animal testing. However, the translation of these methods into regulatory tests (e.g. OECD TGs) that satisfy standard information requirements has been extremely limited. EU-funded research projects invest very little in validation activities since there is little incentive to do so. The main incentive for researchers, in particular academic researchers, is the publication of their results in peer-reviewed journals. The most scientifically rewarding (and publishable) part of a validation study is the demonstration of the relevance of a newly developed method, i.e. to show how the results bring information that is relevant for decision-makers, such as regulators. Similarly, a lot of attention has been given to the development of conceptual testing strategies and demonstration through case studies. This is illustrated by the (draft) position paper produced by the Partnership for the Assessment of Risks from Chemicals (PARC) on validation-related activities [114].

The more costly part of a validation study – and much less popular among (academic) researchers – is the demonstration of the reliability of a method. This requires a significant number of repetitions of the

[112] OECD Principles on Good Laboratory Practice https://doi.org/10.1787/9789264078536-en.

[113] https://joint-research-centre.ec.europa.eu/projects-and-activities/reference-and-measurement/european-union-reference-laboratories/eu-reference-laboratory-alternatives-animal-testing-eurl-ecvam/alternative-methods-toxicity-testing/advisory-and-consultation-bodies/esac-eurl-ecvam-scientific_en.

[114] Partnership for the Assessment of Risks from Chemicals (PARC), 'Validation-related activities within PARC to progress new methods and approaches for hazard and risk assessment of chemicals', produced by PARC WP5 & WP6, Nov 2023 (draft).

experiments, preferably in multiple laboratories (called interlaboratory comparisons or studies, or round-robin studies). Validation studies are being increasingly undertaken by companies (SMEs) that have developed methods that are perceived to have sufficient regulatory relevance, and that have gained support from their national coordinator for the OECD WNT group to submit a project to the OECD to develop a TG. Validation studies have also been carried out by industry associations that want to bring their methods up to the level of a TG, so that they can satisfy regulatory information requirements in their sector. Some years ago, in France, a public-private partnership called 'Pepper' was set up to carry out validation studies (primarily laboratory ring trials to assess reproducibility) on in vitro methods considered relevant for the identification of endocrine disrupting chemicals.

In December 2023, the OECD WNT organised a workshop to discuss operational and financial aspects of validation, which clearly highlighted the organisational complexity and the specialist competences required, and underlined the need for appropriate funding mechanisms taking into consideration the high costs involved (OECD estimates ~EUR 200 000-500 000 for a 'simple' validation of one method) [115].

At present, the validation and regulatory acceptance of non-animal approaches is a long and time-consuming process, which is not considered sufficient for addressing wide-ranging regulatory and societal needs. There is a need to accelerate the availability of validated and regulatory accepted test methods for the safety assessment of chemicals. The European strategy for test method development and validation is a Member-State-born initiative aimed at prioritising and ensuring method development and validation, based on identified regulatory needs for chemical safety assessments and the deficit of validated methods to address those needs, taking into consideration the societal demand for animal-free methods. This initiative will be carried out under the One Substance, One Assessment (OSOA) Expert Group, ensuring a more comprehensive approach to the uptake of alternatives to animal testing.

Standardisation

Standards organisations such as CEN and ISO develop a large variety of standards, some of which are relevant for chemical safety assessments and, to a larger extent, for the assessment of products. Examples include testing standards for medical devices, projects to develop standards for the assessment of nanomaterials, or organ-on-a-chip approaches.

The development of CEN or ISO standards is the result of public-private cooperation involving industry, service providers, public authorities, etc. The European Commission[116] is a major funding body for the development of CEN standards. Each year it publishes an annual work programme setting out its priorities, which may lead to mandates for standardisation.

Standards adopted by CEN or ISO offer some of the advantages offered by OECD TGs. As with OECD TGs, once included in legislation or accepted by regulators, harmonised standards provide industry with certainty on which approaches can be used and how data should be generated. They also facilitate the authorities' assessment of submitted data thanks to the standardisation of the data, and the presumption of conformity with the regulatory requirements covered by harmonised standards the references of which have been published in the Official Journal of the European Union. CEN standards are recognised in the EU and the European Free Trade Association, while ISO standards are intended to be applied worldwide. For this reason, data generated via CEN or ISO standards support the international acceptance of data, which helps to prevent duplication of testing.

At European (CEN) level, 'harmonised standards', support the application of the 'new legislative framework'.

Qualification

[115] OECD workshop report: **Workshop report on operational and financial aspects of validation**

[116] https://single-market-economy.ec.europa.eu/single-market/goods/european-standards/standardisation-policy_en.

‘Qualification’ is another relevant term and concept with a particular meaning in the context of this SWD. It is understood as a voluntary procedure to evaluate a novel method intended for a specific context of use resulting in a ‘qualification opinion’ from an authority.

The EMA offers scientific advice to support the qualification of innovative methods intended for a specific context of use in the research and development of new medicines, including the demonstration of their safety for a market authorisation. The advice is given to applicants by the EMA’s Committee for Medicinal Products for Human Use (CHMP) on the basis of recommendations made by the Scientific Advice Working Party. This qualification process leads to a CHMP *qualification opinion* or CHMP *qualification advice*. On the basis of the qualification advice, the EMA may propose a letter of support, when the novel methodology under evaluation cannot yet be qualified but is shown to be promising on the basis of preliminary data. Qualification opinions and letters are made public on the EMA website. In addition to seeking qualification advice or an opinion, developers of innovative medicines, technologies or methodologies can also engage in dialogue with the EMA’s Innovation Task Force early on in the development process, to help decide on next steps in their development programmes. In recent years, developers and users of alternatives to animal testing have been particularly encouraged to make use of these innovation support measures. Finally, the EMA’s 3Rs (replacement, reduction, refinement) working party is currently supporting the revision of the EMA guideline on the principles of regulatory acceptance of 3Rs testing approaches, which will include the specification of qualification criteria for organ-on-chip based methods (e.g. for addressing safety questions associated with the heart and liver).

Building on EMA experience, EFSA funded a project named NAMs4NANO with the aim of developing a qualification system within its remit to facilitate the regulatory implementation of non-guideline NAMs for a specific context of use (GP/EFSA/MESE/2022/01_Lot 1, see more information in Section 2.2.4 ‘Fostering the use of NAMs for nanomaterials and nanoparticles risk assessment’). The project was outsourced to a consortium of EU Member States and non-EU organisations, with the Bundesinstitut für Risikobewertung (BfR) (German Federal Institute for Risk Assessment) leading this task. A first proposal for a qualification system developed by the NAMs4NANO consortium was published to gather input from the scientific community (Haase et al., 2024). The proposal details a generic framework to speed up the regulatory use of suitable NAMs for chemical risk assessment in food and feed. It includes the establishment of an evaluation process and specific criteria to examine the scientific validity of non-guideline NAMs. The framework is initially proposed for the risk assessment of nanomaterials and nanoparticles, in line with the approach defined by EFSA’s nano guidance. Targeted discussions were held to further improve the qualification proposal and develop an implementation plan, involving agencies and organisations such as the EMA, ECHA, the Scientific Committee on Consumer Safety (SCCS), the JRC, the US Food and Drug Administration, the OECD, the Malta Initiative Board, and the EFSA NanoNetwork, along with scientific and industry representatives. From these exchanges, the need for an infographic to clarify the distinction between validation and qualification was identified. As a result, the figure below was collaboratively produced by the BfR, EFSA, OECD and JRC. All this work will be instrumental in refining the qualification proposal, which is expected to be finalised in 2027.

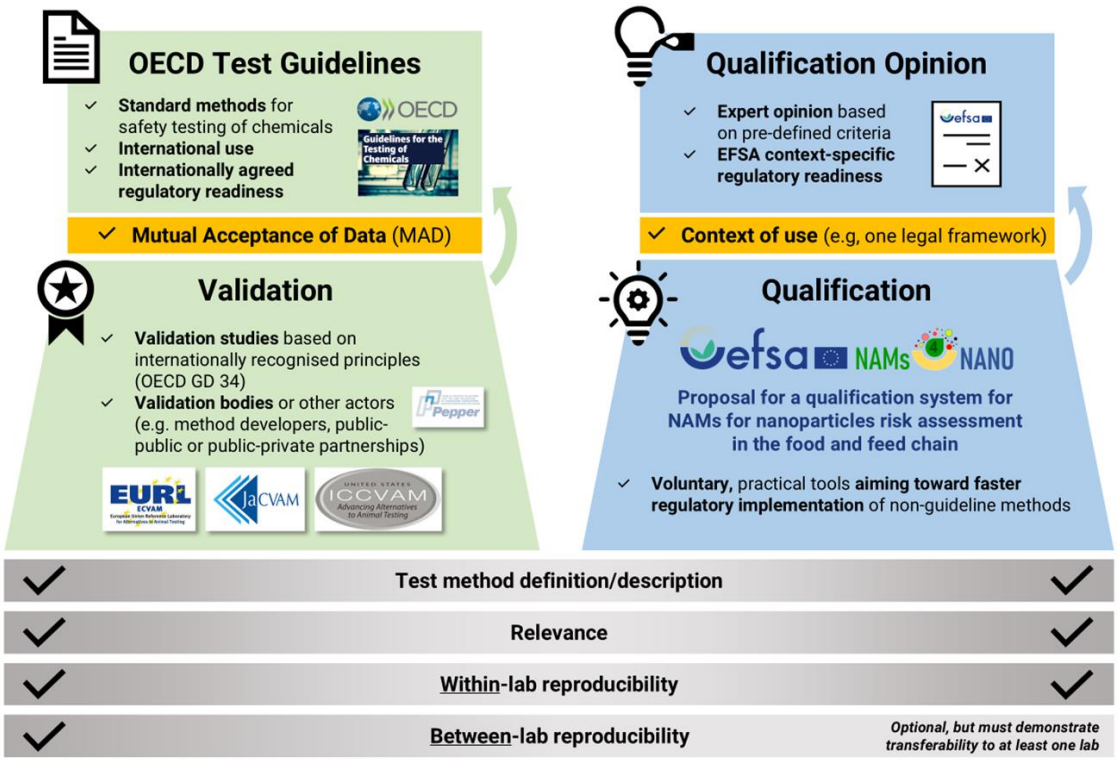


Figure 25: NAMs4NANO graphic analysis of validation vs qualification of alternatives to animal testing. This figure is also available in PDF at: https://www.efsa.europa.eu/it/topics/topic/nanotechnology

For EMA qualification purposes, a 'context of use' is understood to describe the specific intended application of a new approach methodology in medicines development or regulatory assessment, and the conditions under which it can inform regulatory decision-making. Similarly, in the context of the development of an EFSA qualification process, 'context of use' describes the precise circumstances, the intended applications, and the limits under which a NAM could be used for a specific problem formulation (e.g. nanomaterial risk assessment in food and feed).

Agencies may implement their own qualification frameworks to both non-standard methods and test guideline methods to make sure they work in their specific contexts of use. The requirements for context-of-use-based qualification of a TG method may be lower than those for a non-standard method.

OECD TGs include methods used for a particular regulatory purpose; however, this is usually broader than a context of use as applied in EMA and EFSA qualification processes. While a well-defined context of use is central to EMA and EFSA qualification processes, it plays a less critical role in the validation of methods under OECD TGs.

Use of research data for regulatory purposes

Next to data generated with validated or qualified test methods or according to regulatory accepted standards, criteria have recently been developed for the acceptance of research data for regulatory purposes. Examples are the OECD guidance on the use of research data, the pesticides guidance for literature search and the Critical appraisal of evidence for food and feed safety assessments.

Key references and resources

1) Validation, Qualification, and Regulatory Acceptance of New Approach Methodologies - A Report of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). March 2024 (ICCVAM 2024 Report).
2) International Cooperation on Alternative Test Methods, ICATM https://ntp.niehs.nih.gov/whatwestudy/niceatm/iccvam/international-partnerships/icatm https://joint-research-centre.ec.europa.eu/eu-reference-laboratory-alternatives-animal-testing-eurl-ecvam/alternative-methods-toxicity-testing/advisory-and-consultation-bodies/international-cooperation-alternative-test-methods-icatm_en.
3) The EU Reference Laboratory for alternatives to animal testing (EURL ECVAM) https://joint-research-centre.ec.europa.eu/projects-and-activities/reference-and-measurement/european-union-reference-laboratories/eu-reference-laboratory-alternatives-animal-testing-eurl-ecvam_en
4) Tracking system for alternative methods towards regulatory acceptance https://tsar.jrc.ec.europa.eu/.
5) The OECD Test Guidelines ProgrammeGuidelines for the Testing of Chemicals | OECD
6) The OECD IATA Case Studies Project https://www.oecd.org/chemicalsafety/risk-assessment/iata/.
7) OECD guidance document 34 on the validation and international acceptance of new or updated test methods for hazard assessment https://ntp.niehs.nih.gov/sites/default/files/iccvam/suppdocs/feddocs/oecd/oecd-gd34.pdf.
8) A framework for establishing scientific confidence in new approach methodologies. van der Zalm et al. (2022) https://link.springer.com/article/10.1007/s00204-022-03365-4.
9) New approach methodologies workshop: Towards an animal-free regulatory system for industrial chemicals (31 May–1 June 2023) https://echa.europa.eu/-/new-approach-methodologies-workshop-towards-an-animal-free-regulatory-system-for-industrial-chemicals.
10) Building Confidence in New Evidence Streams for Human Health Risk Assessment, National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001;; http://www.nap.edu.
11) Readiness assessment online tools: https://readedtest.u-paris-sciences.fr/ https://scirap.org/Page/Index/ee9102de-4b17-4c3a-86b6-e3e70d6ca3d1/evaluate-reliability-and-relevance.
12) EMA's qualification system: https://www.ema.europa.eu/en/human-regulatory-overview/research-development/scientific-advice-protocol-assistance/qualification-novel-methodologies-medicine-development https://www.efsa.europa.eu/en/news/nanotechnology-promoting-uses-new-assessment-methods
13) NAMs4NANO Proposal for a qualification system within EFSA's remit: https://www.efsa.europa.eu/en/supporting/pub/en-9008.
14) Haase A, Barroso J, Bogni A, Bremer-Hoffmann S, Fessard V, Gutleb AC, Mast J, McVey E, Mertens B, Oomen AG, Ritz V, Serchi T, Siewert K, Stanco D, Usmani SM, Verleysen E, Vincentini O, van der Zande M, Cubadda F, 2024. Proposal for a fit for purpose qualification system for New Approach Methodologies (NAMs) in the food and feed sector. EFSA supporting publication 2024: 21(9):EN-9008. 96 pp. doi:10.2903/sp.efsa.2024.EN-9008.
15) OECD (2025), *Guidance Document on the Generation, Reporting and Use of Research Data for Regulatory Assessments*, OECD Series on Testing and Assessment, No. 417, OECD Publishing, Paris, https://doi.org/10.1787/8d49ec1d-en.
16) Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under Regulation (EC) No 1107/2009, European Food Safety Authority, https://doi.org/10.2903/j.efsa.2011.2092Digital Object Identifier (DOI) (being revised)
17) EFSA Draft guidance on critical appraisal of evidence for food and feed safety assessments https://connect.efsa.europa.eu/RM/s/consultations/publicconsultation2/a0lTk000007cNQz/pc1838

5.6. Safe spaces and regulatory exploration spaces for animal-free methodologies

Context

The European Commission's roadmap towards the phasing out of animal testing calls for innovative regulatory instruments that can accelerate the uptake of animal-free methodologies. Among these instruments, the concepts of a regulatory 'safe space' and a 'regulatory exploration space' (RES) have gained prominence.

Safe spaces refer to platforms that allow applicants to share data confidentially with regulators to discuss the potential acceptability of specific methods for a given case. They provide opportunities for non-binding consultation and exploratory discussions without regulatory consequences, with no obligations on either the submitter or the agency.

Regulatory exploration spaces refer to fora where method developers, stakeholders, academia, NGOs, industry and regulators collaboratively discuss in an informal environment scientific-technical issues related to methods or approaches, their development or improvement, and their application (in different sectors), as well as problem formulations, in order to define new and creative solutions.

A safe space is a bilateral exchange (applicant-regulator) while a regulatory exploration space can be multilateral, including different stakeholders.

The term regulatory sandboxes was not used to avoid confusion with the pharma legislation, which defines a regulatory sandbox as a formal framework authorised by the Commission, which allows innovative regulatory solutions to be tested under supervision and for a limited time. The startup and scaleup strategy also refers to regulatory sandboxes and states that a common EU legal definition will be introduced in the European Innovation Act. In this context, the term RES helps to clearly distinguish the roadmap's informal, exploratory activities from legally formalised regulatory sandboxes.

Stakeholder consultation

Across multiple consultations and events, stakeholders have consistently voiced the need for structured, transparent and trusted mechanisms to support the development and regulatory exploration of animal-free methodologies.

A recurring request concerns the establishment of 'safe spaces': fora for early, non-binding dialogue between industry and regulators. These would allow the confidential discussion of innovative, non-traditional approaches before formal submission, helping to build mutual understanding, clarify regulatory expectations and de-risk innovation. Stakeholders note that existing mechanisms, such as helpdesks or informal exchanges, are not well tailored to the specific challenges of animal-free methodologies and often lack the predictability and strategic value required.

In parallel, many stakeholders have also expressed strong support for the creation of regulatory exploration spaces (RESs). RESs would enable a discussion of novel approaches under regulatory oversight without immediate consequences for authorisation or compliance decisions. Stakeholders consider that RESs could help generate case studies, inform regulatory guidance updates and help identify acceptable use cases, thereby increasing confidence and transparency.

Industry associations in particular have stressed that RESs should reflect real-world conditions, provide legal clarity and support the operational scalability of animal-free methodologies.

Stakeholders further highlight the need for coordination across sectors and agencies, suggesting that bodies such as the EPAA and the EURL ECVAM, and expert coordination groups at ECHA and the EMA could play pivotal roles in convening or supporting these mechanisms.

According to stakeholders, at present, agencies do not operate dedicated RES mechanisms, although existing advisory services and scientific fora could be used to develop and test RES models. The Commission has received consistent feedback that a cross-sector approach to both safe spaces and RESs, possibly under a common EU coordination framework, would be highly beneficial.

Agencies and safe spaces

This section reviews existing support mechanisms at three EU agencies (EFSA, ECHA and the EMA) and explores how the safe-space model could be advanced to foster the regulatory acceptance of animal-free methodologies.

EFSA has developed a comprehensive catalogue of support measures[117] that guide applicants through the entire lifecycle of regulated product applications. These include general pre-submission advice (GPSA), renewal pre-submission advice (RPSA), fast-tracked services for SMEs, ad-hoc clarification meetings and technical hearings, as well as training on the agency's digital submission tools.

EFSA provides advice on dossier preparation and frequently highlights shortcomings in relation to existing guidance. However, it does not engage in discussions on testing strategies; nor does it provide support specific to non-animal testing approaches. While EFSA is not permitted to offer scientific advice directly to individual companies, it may participate in public discussions on broader scientific or regulatory issues. The forthcoming evaluation of EFSA by the European Commission, covering the period from 2017 to 2024 and expected by March 2026, may offer an opportunity to revisit and potentially revise the current approach to pre-submission advice. Collectively, these services boost transparency and procedural clarity and are valued for improving the quality and completeness of dossiers.

ECHA currently offers assistance in minimising the need for animal testing or applying animal-free methodologies via the testing proposal evaluation process. ECHA also offers general regulatory assistance through its Helpdesk. The high volume of submissions managed by ECHA limits opportunities for additional informal dialogue of an exploratory nature.

The EMA has the most developed structure for fostering early, iterative dialogue on innovative new approach methodologies. In this context the term 'new approach methodology' is used by the EMA to refer to methods that can not only replace but also reduce animal testing. The EMA's Innovation Task Force provides a confidential forum for early-stage discussions on novel technologies, including new approach methodologies. Additionally, portfolio and technology meetings support strategic engagement on large product pipelines, offering opportunities to discuss innovative approaches, including new approach methodologies. Through the scientific advice mechanism at the EMA, applicants can obtain formal guidance on medicinal product development, including the use of weight-of-evidence approaches that incorporate new approach methodologies Specifically, for regulatory acceptance of a new approach methodology, applicants may request qualification advice, which offers regulatory feedback on the method's suitability for use in product development within a defined context of use. The EMA has also set up a 'safe harbour' pathway for the voluntary submission of new-approach-methodology data, which is used solely for the purpose of collecting information on new approach methodologies for possible future regulatory acceptance with no consequences for regulatory decision-making on medicinal products. This process is under review to further enhance the safe-space approach and facilitate data sharing.

[117] https://efsa.onlinelibrary.wiley.com/doi/10.2903/sp.efsa.2021.EN-6472

Future steps for safe spaces

In both EFSA and ECHA there is no clearly defined safe-space mechanism for the early consideration and discussion of non-traditional approaches. Nor is there a dedicated structure in place to support their systematic integration into regulatory processes. While existing support measures have their uses, they do not constitute a true regulatory safe space as envisioned in this document. All advice must remain strictly within the boundaries of existing legislation and guidance.

An additional layer of complexity stems from the requirement that all initiatives must operate in full compliance with sector-specific legislation and the broader governance and legal framework. Within these constraints, however, it remains unclear how more flexible and innovative processes, capable of functioning as genuine safe spaces, can be introduced.

To explore the feasibility and practical implications of establishing such mechanisms, it is recommended that stakeholders, the Commission and the agencies discuss possible ways of implementing safe spaces. Further discussions on practical implementation should help ensure that future actions are grounded in evidence and informed by real-world experience.

It is therefore proposed to organise a workshop, inviting stakeholders, Commission services and the agencies to discuss potential ways of implementing safe-space mechanisms and to identify practical challenges and enablers. The workshop could also be used to exchange views on the EMA's ongoing safe-space experience, which may provide useful insights for similar approaches at other agencies, such as ECHA and EFSA.

Depending on the outcome of these discussions, and if a need is identified, a continued dialogue on practical implementation could be pursued before committing to any large-scale initiative requiring significant resources. This approach would ensure that future actions were evidence-based and informed by real-world experience.

Given the level of interest in this area, the workshop and any potential follow-up activity should be conducted transparently and in close cooperation with relevant stakeholders. The EPAA could take the lead in organising the workshop, providing the necessary flexibility in both design and implementation, and involving agencies, stakeholders and Commission services.

The Commission services will assess the outcomes of the workshop and subsequent discussions. On the basis of the lessons learned, the roadmap steering team could then determine whether further legal adaptations or additional resources were needed to scale up the safe-space initiative effectively.

The regulatory exploration space concept

The concept of a RES has emerged as a promising tool to support the structured implementation and evaluation of innovative methodologies, particularly animal-free methodologies, in real-world regulatory contexts. In the context of chemical safety assessments, RESs could enable method developers, regulators, industry, academia and civil society organisations to work collaboratively in refining problem formulations, sharing relevant data, and assessing the performance and suitability of animal-free approaches before these are codified into regulatory guidance or legal frameworks. This would support mutual learning, de-risk innovation and accelerate confidence-building in new methodologies. A number of key principles underpin the effective operation of a RES. These include the need for clear governance rules, predefined criteria for evaluating outcomes, appropriate protections for confidential data, and mechanisms to ensure transparency and the dissemination of the insights gained. Embedding such principles in the operating procedures of competent authorities would help

foster regulatory trust, reduce uncertainty for innovators and contribute to the EU's broader ambition to lead in global non-animal safety science.

Although the concept has not yet been fully formalised across EU regulatory agencies, several existing initiatives already display key features of RESs and can serve as practical models for further development. At ECHA, the Persistent, Bioaccumulative and Toxic Expert Group operates as an informal advisory body, offering scientific input on complex cases involving new or borderline methodologies. Although not a formal RES, it provides a protected space for early consideration of innovative scientific approaches. In addition, ECHA's Expert Group on Endocrine Disruptors offers a dedicated forum for early, non-binding discussions on evolving methodologies relevant to endocrine disruption. Finally, the Alternative Methods Team is also involved in informal consultations, providing feedback on industry proposals and scientific advice to ECHA processes, Marie Skłodowska-Curie Actions and industry. These informal groups enable regulators and stakeholders to explore complex scientific issues in a structured yet flexible manner, further reinforcing ECHA's capacity to support RES-like engagement. EFSA, while not currently operating a formal RES, has developed stakeholder engagement platforms that serve many of the same purposes. Through public consultations, expert colloquia and technical stakeholder dialogues, EFSA provides structured opportunities for stakeholders to engage in the early discussion and exploration of new scientific approaches. These platforms enable multi-directional learning and allow for the transparent exchange of knowledge in a non-binding context, supporting the iterative development of regulatory thinking and scientific guidance.

Other examples of similar activities are the Scientific Committee on Consumer Safety (SCCS) and the EPAA non-animal method (NAM) Designathon.

The SCCS, operating under the Cosmetics Regulation that fully banned animal testing in the cosmetics sector, has implemented a dual-submission practice whereby both traditional and NAM-based data can be submitted for evaluation. This framework enables the informal consideration of NAMs alongside established methods and serves as an effective RES for next-generation risk assessment (NGRA) strategies.

The EPAA NAM Designathon, for example, has established a structured co-creative setting where regulators, scientists, industry representatives and NGOs jointly develop and assess NAM-based testing strategies using real-world case studies. These sessions enable structured experimentation in a non-binding format, promoting regulatory familiarity with novel approaches and facilitating alignment on scientific expectations.

ECHA and EFSA are also active partners in the Accelerating the Pace of Chemical Risk Assessment (APCRA). The APCRA initiative serves as a practical example of a RES-like setting at international level. It enables regulators from multiple countries to collaboratively explore and evaluate the use of animal-free methodologies in chemical risk assessments through case studies, without regulatory consequences. APCRA demonstrates how structured, non-binding collaboration can build confidence in innovative approaches and support regulatory convergence.

Future steps for RES

While the agencies have not yet fully formalised RES mechanisms, stakeholders have called for them to be set up and coordinated under a common EU framework. Existing initiatives by platforms such as the EPAA, including the NAM Designathon and User Forum, demonstrate the value of such collaborative settings. Building on these models, there is a clear recommendation to expand and institutionalise regulatory RESs across sectors, enabling systematic, low-risk testing of innovative approaches and fostering convergence among EU agencies and different legislative sectors.

The EPAA is well positioned to support the facilitation or coordination of such initiatives. Existing formats, including the EPAA Designathon and User Forum, have demonstrated the feasibility of multi-stakeholder collaboration and could serve as the basis for more structured RES activities. Future

development should aim to institutionalise RESs across sectors, ensuring that agencies can test, adapt and build confidence in new methodologies in a controlled, transparent and legally secure manner.

EFSA, taking into account the EMA's experience, could explore the possible establishment of a RES for the qualification of non-animal approaches, with nanomaterials risk assessment as a first example of possible implementation.

For maximum impact, RES activities should be designed to complement existing initiatives and to enable coordination across agencies, Member States and legislative frameworks. A networked model, linking agency-specific structures on animal-free methods with existing advisory processes, would foster a coherent European ecosystem for innovation in chemical safety assessments.

In conclusion, the development of RESs represents a necessary next step in supporting the transition to animal-free safety assessments. Such structures will enable the controlled and practical evaluation of innovative methods, promote cross-sector learning and provide the necessary evidence base for future guidance updates and regulatory reform.

Conclusion and forward look

RESs in synergy with safe spaces would strengthen the EU's leadership in evidence-based, non-animal science.

Setting up such mechanisms, anchored by collaborative structures that coordinate scientific evaluation, stakeholder dialogue and cross-sector learning, would respond directly to recommendations emerging from the roadmap. It would also position the EU as a global leader in innovative, animal-free safety assessments.

All the different activities will be coordinated by the roadmap steering team, which will monitor progress and ensure alignment with wider EU policy objectives.

6. Indicators – managing change by measuring it

Rationale and background

In the context of the Commission's roadmap, there is a need to establish indicators to monitor progress towards the final goal of phasing out animal testing in chemical safety assessments. The aim is to develop a set of indicators that can be used to develop both cross-sector and sector-specific narratives. It is possible that some indicators will be applicable to all sectors, and that others will be sector-specific, depending on the nature of the available underpinning data. It is also expected that there will be methodological differences between sectors in terms of data availability and accessibility. Achieving a better overview could be the basis for future efforts to harmonise data collection. It is also considered that certain indicators might be more informative in the short and medium term, and others in the long term. In addition, when we start off with a first set of indicators to measure progress to better manage change, we might decide to add others or revise some in the first set to gain better knowledge on progress during implementation. Some indicators might also cease to be relevant if progress to an intermediate goal has been achieved, e.g. testing in a second species is no longer conducted.

Crowd sourcing ideas from stakeholders

During the Animal-Free Chemicals Safety Assessment Conference, 4-6 March 2025, co-organised by the EPAA and the Commission, two parallel break-out groups were set up to discuss indicators and how to best measure progress during implementation of the roadmap. About 40 participants representing different stakeholders took part in these discussions, and there was also a reporting back session in plenary, collecting additional input from the full audience. The questions discussed were:

- What does success look like to you?
- What change is necessary to reach the successful outcome?
- Who should host and report on indicators?

Many indicators were suggested, and after the conference a small core team made up of representatives of the Commission, industry and an animal welfare NGO worked further on the input and circulated a table of proposed indicators to the break-out groups' participants for comments. The resulting list of indicators was presented at the third workshop to discuss the roadmap to phase out animal testing for chemical safety assessments, 16-17 June 2025 and was further updated based on the feedback. The initial and broad collection of ideas from all stakeholders that will be further evaluated for relevance and feasibility is included in Annex 1.

The indicators were classified in four categories to make sense of the underlying types of change:

1. engagement indicators
2. status indicators
3. counter-progress (weakness) indicators
4. progress (3Rs) indicators.

Engagement indicators aim to capture public awareness and support, knowledge sharing, collaborations and partnerships. Status indicators are related to development and validation or qualification of non-animal or alternative approaches, funding and investment, cost-effectiveness, industry adoption, and regulatory acceptance. The counter-progress or weakness indicators aim to identify barriers to acceptance and regulatory uptake, efficiency and reliability. Progress indicators are directly related to the ultimate aim of the roadmap, i.e. the phasing out of animal testing.

Composite indicators, i.e. several indicators used in combination, are defined as a separate group.

Usefulness of indicators

Indicators should be relevant, reliable and feasible in order to measure change, understand progress and inform priorities. It is important to understand that an indicator is not only a number to be used in debates or to give the appearance of success. If used carefully and impartially, indicators will serve as a means to better monitor the transition to animal-free chemical safety assessments and to direct roadmap actions. For example, are the activities supported by the roadmap steering team and set out in the transitional initiatives leading to the expected progress? The indicators might tell a different story from that intended by those helping to implement the roadmap. For instance, the number of alternatives to animal testing developed does not have any impact if it is not regulatory relevant. Even though it was the intention for alternatives to animal testing to be used for regulatory testing, it might not be suitable for that purpose in the end. Cost-benefit considerations need to be measured to gain understanding in how to improve the ratio of alternatives to animal testing that can be used for regulatory testing and make it work in favour of progress.

Indicator description

To make further progress with the description of an indicator, we suggest that the following information must be available and identified for each indicator:

1) short title of indicator
2) owner
3) definition
4) expected usefulness
5) data source(s) including their availability/accessibility
6) methodology for finding, processing and interpreting the data
7) limitations in the interpretation of the indicator
8) frequency and timing of reporting.

First suggested prototype indicators

We have completed the indicator description for the indicators listed below (Annex 2, Example indicators) from the table in Annex 1 and suggest that these can be included in the first set of prototype indicators to be evaluated in a trial phase.

Type of indicator	First set of prototypes to evaluate
Engagement	Awareness of non-animal approaches and their possible application in chemical safety assessments (based on interviews with stakeholders)
	Use of positive and negative language in public discourse, e.g. political speech, corporate speeches and publications (based on publicly available material)
Status	Follow-up of roadmap actions to assess the actionability and continued relevance of actions (in terms of inputs and relevant outputs)
	Relative cost of a chemical safety assessment based on non-animal approaches compared with the traditional animal test (both approaches providing equivalent information)
Counter-progress	Use of non-animal approaches to trigger additional in vivo testing (according to regulatory information requirements and guidance)
	Bias in research funding towards the use of animal tests over non-animal approaches (number of grants fostering non-animal studies)
Progress	Phasing out of animal testing fulfilling EU regulatory information requirements (number of animals used to fulfil regulatory information requirements, horizontal and per legislative sector)
	Reduction of severity in animal procedures where these tests are still performed (based on severity, horizontal and per-legislative-sector scores)

Hosting an indicator and tell a story

It is envisaged that the set of indicators applied during roadmap implementation will not be fixed over time and will have distributed ownership (i.e. responsibility for developing and reporting). The Commission might not be the right developer and host for all indicators. For example, the EU agencies already report indicators that are relevant for the roadmap. Additional indicators are listed in the table in Annex 1, some of which other stakeholders might have an interest in refining and reporting. An indicator can be sector specific and – even when only reported by a single company or organisation – still be informative for a better understanding of how to tackle different issues to achieve progress. Therefore, specific stories might be of great interest and value, while the more horizontal indicators will tell a more general story.

It is recommended that the roadmap steering team take on the task of evaluating selected indicators for their usefulness during the roadmap's implementation.

Next steps

The Commission Communication sets out a few indicators, while the Commission may identify a longer and more detailed set of indicators when the implementation phase starts.

References

EPAA (2025) Animal-Free Chemical Safety Assessment Conference, Flash Report, https://webgate.ec.europa.eu/circabc-ewpp/d/d/workspace/SpacesStore/4015d661-04da-440e-96d9-a937e955d341/download.

Conference Report, d7994cf5-4b38-9a8a-9cbc-0c89da0dcad8.

EC (2025) Third workshop to discuss the roadmap to phase out animal testing for chemical safety assessments, 16-17 June 2025, Helsinki, hosted by ECHA, https://echa.europa.eu/pt/-/third-workshop-to-discuss-the-roadmap-to-phase-out-animal-testing-for-chemical-safety-assessments.

EC (2024) ALURES – ANIMAL USE REPORTING - EU SYSTEM, EU statistics database on the use of animals for scientific purposes under Directive 2010/63/EU. ALURES statistics - Section 1 : Numbers of animals used for research, testing, routine production and education and training purposes - EU statistics database on the use of animals for scientific purposes

Indicators - Annex 1

Initial and broad collection of ideas from all stakeholders that will be further evaluated for relevance and feasibility. The eight prototype examples mentioned in the text and further described in Annex 2, are highlighted in green.

1. ENGAGEMENT INDICATORS (Public Awareness and Support, Knowledge Sharing, Collaborations and Partnerships)				
Title	**Definition**	**Must be relative to**	**Data holders**	**Comments**
Create confidence in applying new testing strategies through bilateral discussion *(Example 1)*	Number of bilateral meetings between companies and agencies about testing strategies leading to a reduction of animals used or involving non-animal approaches	Each agency and substance	Agencies	Transparent meeting records to be made available, in general without any substance specific details
Create confidence through communication to better guide dossier submitters towards possible acceptance of information from non-animal approaches	Number of communications providing feedback from agencies to dossier submitters, made publicly available	Each agency and substance	Agencies	The records need to be transparent but can omit business-sensitive or confidential information.
Knowledge Sharing supporting collective development towards more successful use of non-animal approaches	Number of case studies on implementation of non-animal approaches with regulatory relevance and shared with broader audience	Sector and context	Agencies, Research projects, Industrial associations	Sharing both successful and unsuccessful application of non-animal approaches also as a summary of information and avoiding any confidential information
Awareness of non-animal approaches and their application in chemical safety assessment	Based on stakeholder interviews regarding acceptance of non-animal approaches in chemicals safety assessment (focusing on stakeholders not contributing directly to chemicals safety assessment)	Description of stakeholder interviewed and legal area(s) in scope of the roadmap	Researchers performing interviews	Evaluate barriers for acceptance phasing out of animal testing, e.g. lack of training, upskilling, fear of losing jobs.
Use of positive and negative language in public discourse, e.g. political speech, corporate speeches and publications *(Example 2)*	Sentiment analysis of specific communication, like comments from relevant regulatory committees, official communications and consultations by the Commission, MS authorities and other stakeholders	Specific communicative contexts	Publicly available	This could be made also retrospectively to observe current trends and the development during the roadmap implementation.
Recognitions awarding faster progress	Number of awards for accomplishments related to the phasing out of animal testing in chemical safety assessment	Description of context, frequency and popularity	Collected by 3R centres, Member State authorities, industry associations, NGOs	This could be made also retrospectively to observe current trends and the development during the roadmap implementation.

2. STATUS INDICATORS (Development and Validation of Alternatives, Funding and Investment, Cost-effectiveness, Industry Adoption, Regulatory Acceptance)				
Title	**Definition**	**Must be relative to**	**Data holders**	**Comments**
Follow-up of roadmap actions to assess the actionability and continued relevance of actions *(Example 3)*	Number of outputs related to each action presented in the Communication or Staff Working Document	Each action	Various - to be specified for each action	Over time, it is expected that roadmap actions will be refined, added, or discontinued, according to progress.
Maintain innovation and build confidence by establishing non-animal methodology batteries	Number of new versions of non-animal methodology batteries accepted to ensure protection levels	Per substance and decisions made	Industry	Not as relevant for R&D pipeline screening methods, but rather regulatory application
Build capacity by dedicating a higher percentage of staff to animal-free safety assessment	Number of FTEs working on issues related to animal-free and animal-based safety assessment in different sectors	Total number of FTEs working on issues related to chemical safety assessment and sector	Agencies, industry, CROs	The combined information provides insight in whether the non-animal approaches are supporting the phasing out of animal use or is used for enhanced understanding of the animal studies.
Maximise cost-benefit when replacing animal methods with in vitro testing	Cost of in vitro tests carried out for safety assessment compared to the cost of replaced animal methods	Per substance assessed for regulatory safety assessment	Agencies, industry, CROs	Other information might also be considered in assessment of a substance, but this indicator focuses on in vitro testing.
Accelerate validation and acceptance of non-animal approaches	Number of new non-animal approaches with successful validation acc. to OECD GD34and publication within the test guidelines programme or included in case study publications (IATA case studies project)	Time discussed at the OECD prior to acceptance or estimate of the research funding allocated per non-animal approach	OECD	To better understand how to make the process more efficient
Prioritise the development of new non-animal approaches to fulfil a higher percentage of regulatory information requirements	% of regulatory standard information requirements for which OECD validated test methods and defined approaches are available	Time discussed at the OECD prior to acceptance or estimate of the research funding allocated per non-animal approach	Agencies	Similar to the one above but is relative to the breadth of regulatory needs. Non-animal approaches for a specific regulatory information requirement typical include several test methods
Prioritise development of non-animal approaches to cover a higher number of chemicals and a higher number of endpoints	% of chemicals and endpoints assessed in a regulatory framework relying on non-animal approaches and without new animal testing	Time discussed at the OECD prior to acceptance or estimate of the research funding allocated per non-animal approach	Agencies	Could also be considered whether there is a time frame for legal uptake within the EU

Accelerate progress within the EU by regulatory acceptance prior to acceptance at OECD level, and encourage progress within other jurisdictions	Number of non-animal approaches included in EU guidance and regulation prior to adoption by the OECD replacing animal testing	Number of standard information requirements/ endpoints for which animal testing is required, or research funding and cost/ time of validation allocated per non-animal approach.	Agencies	Information on promotion of acceptance outside the EU could be collected.
Use of non-animal research data to build confidence and accelerate acceptance of non-animal approaches	Frequency of use of non-standard non-animal research data in regulatory assessments	Per endpoint or per substance	Agencies, Research authors, publishers	OECD guidance on better use of research data in regulatory assessments
Resource allocation for confidence building and capability development	Funding provided for case studies, trainings and workshops aiming at confidence building in non-animal approaches	Per use-sector and number of relevant chemistries	Commission, MS authorities, agencies, industry	
Resource allocation for development of non-animal approaches to accelerate regulatory acceptance of non-animal safety assessment	Funding provided for development non-animal approaches that are eventually accepted for regulatory implementation	Per use-sector, or percentage of overall research funding on NAM development	Commission, MS authorities, agencies, industry	
Development of guidance on reviewing dossiers including data from non-animal approaches	Cost and time for reviewing non-animal approach-based dossiers or realistic case studies	Cost and time used to review corresponding dossiers including *in vivo* studies	Agencies	Consider EPA metrics https://www.epa.gov/chemical-research/new-approach-methods-nams-animal-use-metrics-research-and-development. Communication for regulators is important to help incentivise NAM use.
Relative cost of a chemical safety assessment based on non-animal approaches compared with the traditional animal test (both approaches providing equivalent information) *(Example 4)*	Market price for running a substance through a non-animal battery/defined approach under GLP and market price for the corresponding animal test	Per chemical and endpoint	CROs, industry	Needed: guidance on what is included and excluded in the cost calculation Includes invertebrate testing when replacing vertebrate testing.
Affordability of manufacturers to replace animal testing with non-animal approaches	Manufacturers' Willingness to Pay (WTP) for replacement of *in vivo* testing estimated through surveys or by observing real market choices	Cost per endpoint and per sector	Industry, consultants	Is there interest from manufactures to start ethical labelling on all consumer products?
Affordability of consumers NAM-based to replace animal testing with non-animal approaches	Consumers' Willingness to Pay (WTP) for ethical/cruelty-free products through surveys or by observing real market choices	Cost per product and per sector	Consumers	To inform the consumers, the cruelty-free products need to be labeled as such, which could be of marketing interest to do also outside the cosmetics sector.

Influence of consumer behaviour on the market	Market share of ethical/cruelty-free products	Other products in the same sector	Industry	To inform the consumers, the cruelty-free products need to be labeled as such, which could be of marketing interest to do also outside the cosmetics sector.
Competitiveness of CROs and start-ups providing non-animal solutions	Added value (euros, % of GDP) generated by non-animal services sector compared to the loss of GDP in animal services sector	Other sectors, 3rd country regions	MS authorities	Link to EU start-up and scale-up strategy
Higher number of start-ups providing non-animal solutions	Amount of venture capital given to start-ups specialising in non-animal services	Other sectors	MS authorities	Link to EU start-up and scale-up strategy
Adaptations/ derogations to animal testing mandated by legislation	Number of adaptations allowed to avoid animal testing when fulfilling information requirements	Per substance and information requirement	Agencies	US EPA collects metrics on waivers applied with monthly meetings to discuss and update this data input. Could this approach work for EU agencies?
Regulatory guidance updates within the EU	Number of revisions of guidance per year enabling phasing out of animal testing	Compare with baseline from publication of the roadmap and mapped to the causes triggering the update	Agencies	Could be done for each of the 15 identified legal sectors in the scope of the roadmap.
Legislative updates within the EU	Number of legal text revisions enabling phasing out of animal testing	Compare with baseline from publication of the roadmap and mapped to the causes triggering the update	Commission	Could be done for each of the 15 identified legal sectors in the scope of the roadmap.
Acceptance of non-animal approaches for regulatory submission (including use of invertebrates when replacing vertebrates)	Number of endpoints covered solely by non-animal data in submissions	Number of submissions AND reduction of animals used	Agencies	Might be difficult for pharmaceuticals. The 117(3) report from ECHA provides a good summary under REACH, can this be replicated for other agencies?
Hazard classification based on non-animal approaches (including use of invertebrates when replacing vertebrates)	Number of CLP hazard classes for which classification relies on non-animal approaches	Total number of human health and environmental CLP hazard classes, relevant to animal testing	Commission, Agencies	
Risk assessment based on non-animal approaches (including use of invertebrates when replacing vertebrates)	Number of substances/endpoints for which hazard characterisation relies entirely on non-animal approaches (REACH, PPPR, BPR, any other relevant piece of legislation)	Total number of risk assessed substances/endpoints, including updated assessments	Commission, Agencies	

Enable innovation despite animal testing ban for cosmetics	Number of new cosmetic ingredients put on the market	Compare with baseline since ban (2013)	MSs, Commission, Industry	Need to differentiate between ingredients which need listing and others?
Protection of human health and the environment while accelerating progress towards phasing out of animal testing	Number of hazardous chemicals per endpoint with risk management decisions based on evidence from non-animal data	Compare with baseline since publication of the roadmap		How to prove that the decision was correct, not protective or over-protective? On the other hand, how do we know whether we are doing the correct decisions based on animal testing?
Maintain or improve ability to innovate chemicals (REACH)	Number of new chemicals registered and/or cost for registration dataset	Compare with baseline since publication of the roadmap. Total number and/or cost of registered chemicals during certain time period	ECHA	Number of new chemicals will also depend on many other factors such as global economy, energy cost, etc
Phasing out of second species requirements	Number of information requirements for second species deleted in legislation or regulatory guidance	Number of data requirements for second species at baseline (publication of the roadmap)	Agencies	Phase out the less appropriate species…
Non-animal approaches used in self-classification	% of self-classifications based (partially or entirely) on non-animal approaches under CLP as reported by ECHA	% of harmonised classified substances	ECHA	
Tracking short, medium and long-term opportunities for replacement of animal tests	Broad scope tracking of all regulatory tests on animals in scope of the roadmap which can/can't be phased out	Individual animal tests, performed across all legislation in scope of the roadmap	Commission and agencies	Changes in the population of short/medium/long-term "baskets" will illustrate progress. An important component is the collection of information and analysis of why animal testing persists.

3. COUNTER-PROGRESS (WEAKNESS) INDICATORS

(Barriers to Adoption, Efficiency and Reliability)

Title	Definition	Must be relative to	Data holders	Comments
Use of non-animal approaches to trigger additional *in vivo* testing *(Example 5)*	Number of animal tests carried out for endpoints triggered by the outcome from a non-animal approach in regulatory information requirements and guidance	Number of substances assessed for these endpoints	Agencies	
Encouraging more robust read-across arguments provided by submitter and more confidence in accepting read-across arguments	Number of read-across arguments for specific information requirements not accepted or rejected	Total number of substances with information submitted for the specific endpoint	Agencies, SCCS	Also include lessons learned

Tracking final, full replacement of animal tests	Identification and tracking of the number of animal tests (conducted per year) which continue to be conducted despite appropriate non-animal approaches being broadly available	The specific animal test in question, across all EU legislation in scope of the roadmap	Agencies	This would highlight *very near term* opportunities where the barriers to full replacement are relatively easily surmountable. An important component is the collection of information and analysis of why animal testing persists
Bias in research funding towards the use of animal tests over non-animal approaches *(Example 6)*	Number of grants fostering non-animal studies and number of grants fostering *in vivo* studies	Total number of grants	Research funders	

4. PROGRESS (THREE Rs) INDICATORS

Title	Definition	Must be relative to	Data holders	Comments
Phasing out animal testing fulfilling EU regulatory information requirement *(Example 7)*	Number of animals used for regulatory studies for regulatory studies, horizontal and per legislative sector	To a baseline, also set prior to the publication of the roadmap when data is available	ALURES (EC, 2024) – testing performed within the EU, Industry/Agencies – testing performed outside the EU	REACH A.117 report on alternatives from ECHA has previously provided useful data on numbers per OECD TG. Useful to indicate whether a submission is fulfilling more than one Information Requirement in EU legislation.
Phasing out animal testing carried out within the EU fulfilling **non-EU** regulatory information requirement only	Number of animals used for regulatory studies fulfilling non-EU requirements	To a baseline, also set prior to the publication of the roadmap if data is available	ALURES, industry	Specify other jurisdictions requiring the testing
Reduction of severity in animal procedures where these tests are still performed *(Example 8)*	Severity scores, horizontal and per legislative sector	Study type and endpoint including studies made to fulfil the same test guideline	ALURES	Perhaps possible to map the studies currently used in safety assessment with highest severity, and then see whether they reduce, perhaps looking to a few specific regulatory uses
Waiving of animal studies based on REACH, Annex XI provisions and gaining confidence in non-animal approaches	Number and volume of chemicals with data requirements being fulfilled by Annex XI adaptations	Total number and volume of chemicals registered	ECHA	
Waiving of animal studies based on exposure under REACH (requires Annex XI/guidance update)	Number and volume of chemicals with data requirements being fulfilled by exposure-based waiving	Total number and volume of chemicals registered	ECHA	

Use of a non-animal approach for a certain substance which is possible to apply to other substances	Number of pharmaceutical ingredients to which the same assessment approach was applied	Total number of pharmaceutical ingredients assessed for the same endpoint	EMA	
Waiving of animal studies in veterinary medicines	Number of waivers of animal tests accepted in vaccine batch testing	Total number of waiver requests	EMA, Industry, AnimalHealthEurope	
EU impact the phasing out of animal testing internationally	Number of times when an EU initiative succeeded in phasing out an animal test in OECD guidance or replace GHS criteria based on animal testing etc	Total number of EU initiatives to phase out animal testing in international context	Commission	Useful to also describe the initiatives

COMPOSITE INDICATORS

Title	Definition	Must be relative to	Data holders	Comments
Summary of multiple indicators providing a clearer information	A composite indicator is a useful tool for policy making and public communications in conveying information on performance	Any information defined for the individual indicators being components of the composite indicator	JRC	The Competence Centre on Composite Indicators and Scoreboards at JRC could help to develop composite indicators that integrate and summarise multiple indicators.

Annex 2 – Example Indicators

Example 1

Awareness of stakeholders of non-animal approaches used/possible to use in chemicals safety assessment

Owner

Researchers performing interviews (possibly tasked by the JRC)

Definition

Stakeholders' acceptance of non-animal approaches in chemicals safety assessment covering legislation under the 15 legal areas in scope of the roadmap (focusing on stakeholders not contributing directly to chemicals safety assessment)

Expected utility

Understanding barriers and perceptions of changing from animal to non-animal approaches, and thereby inform confidence building among these stakeholder groups

Data source(s) including their availability/accessibility

Interviews performed by researchers with different stakeholders, summary outcome made publicly available

Methodology for finding, processing and interpreting the data

The interviews will be carried out possibly by the same researchers, and in any case, following the same methodology and questionnaire on each occasion. If considered informative, some questions can be added to better understand development. Representatives of the same stakeholder groups must be selected for each occasion. It might be considered whether desired and feasible to interview additional stakeholders, if there is evidence that the coverage is not complete.

Limitations in the interpretation of the indicator

The number of interviews that can be performed, the selection of stakeholders. Risk for variation in methodology and interpretation of results, if carried out by different researchers. The indicator is qualitative rather than quantitative and therefore might be less specific when measuring change. However, it can be considered to rank the replies to the interviews, in for example five groups, by the researched, which would give some means of comparing interviews, as well as results from different years.

Reporting (frequency and timing)

Starting 2026, and to be repeated every 2 years.

Example 2

Short title of indicator

Use of positive and negative language in public discourse, e.g. political speech, corporate speeches and publications

Owner

Researcher collecting information (possibly tasked by the JRC)

Definition

Sentiment analysis or critical discourse analysis of specific communication, like comments from relevant regulatory committees, official communications and consultations by the Commission, MS authorities and other stakeholders

Expected utility

Understanding barriers and perceptions of changing from animal to non-animal approaches, and thereby improve communication and provide information mor efficiently

Data source(s) including their availability/accessibility

Interrogation of publicly available information in scope of the definition

Methodology for finding, processing and interpreting the data

AI informed research of publicly available information sources, evaluation of relevant content and classification in for example in five groups: (1) positive to phasing out of animal testing, (2) positive to phasing out of animal testing, but reluctant to feasibility, (3) neutral, (4) negative to phasing out of animal testing, but only in specific fields of application, (5) negative to phasing out of animal testing. If possible, specify context and owner of statements, to investigate how this information can be informative for any additional conclusions or suggested follow-up actions. This could be supplemented in cases with an in-depth analysis of specific texts using critical discourse analysis.

Limitations in the interpretation of the indicator

Misinterpretation of statement when taken out of its context. Discourse analysis is less subject to this bias.

Reporting (frequency and timing)

Starting retrospectively, for example 2020, and to be repeated every year.

Example 3

Short title of indicator

Follow-up of roadmap actions to assess the actionability and continued relevance of actions

Owner

Various - to be specified for each action

Definition

Number of outputs related to each action presented in the Communication or Staff Working Document

Expected utility

Over time, it is expected that roadmap actions will be refined, added, or discontinued, according to progress.

Data source(s) including their availability/accessibility

Collected by the Commission, and to be evaluated by the Roadmap Steering Team. The outputs are assumed to be publicly available, reported in context of specific projects, for example through the EPAA, or any registered transitional initiative.

Methodology for finding, processing and interpreting the data

The Commission with support of the Roadmap Steering Team, collect all relevant outputs to the roadmap actions, which is essential for evaluating progress towards phasing out of animal testing. The expected outputs will be mapped against the actions, and made publicly available, to allow for any stakeholder to complete the information.

Limitations in the interpretation of the indicator

The indicator is reporting on quantity rather than quality, and it could be evaluated whether it would be possible to include some qualitative element, which even could allow for a quality appraisal, to present together with the quantitative indicator.

Reporting (frequency and timing)

Starting 2027 and to be repeated every year

Example 4

Short title of indicator

Relative cost of a chemical safety assessment based on non-animal approaches compared with the traditional animal test

Owner

CROs, industry

Definition

Market price for running a substance through a non-animal battery/defined approach under GLP and market price for the corresponding animal test

Expected utility

Over time, the non-animal approach alternative might become more affordable, due to e.g. legal requirements and upscaled testing compared to the current animal testing for a specific endpoint. By following trends, from first introduction of new non-animal approaches to a broader use, we will learn how to tackle both increments in cost for new specific information requirements and how to efficiently lower overall costs (through smart testing strategies).

Data source(s) including their availability/accessibility

Data to be collected by Industrial Associations or directly from CROs on possible replacement approaches.

Methodology for finding, processing and interpreting the data

The Commission with support of the Roadmap Steering Team to list all animal methods in scope of the roadmap that could potentially be replaced by non-animal approaches and publish this list. The Commission with support of, for example, EPAA, collect data on market prices both for animal methods and the non-animal approaches providing equivalent information. For those already in use in the EU, also retrospective information might be collected.

Limitations in the interpretation of the indicator

Variation in market prices (including fluctuations in exchange rates) for testing not performed within the EU, and difficulties to understand how much of the testing is done outside the EU, and what the real change in cost is.

Reporting (frequency and timing)

Starting 2027 and to be repeated every year

Example 5

Short title of indicator

Use of non-animal approaches to trigger additional *in vivo* testing

Owner

Agencies

Definition

Number of animal tests carried out for endpoints triggered by the outcome from a non-animal approach in regulatory information requirements and guidance

Expected utility

The aim is to avoid triggering animal testing, and to enable a decision based on the information obtained from non-animal approaches. This can be achieved by better testing strategies, or enhancement of non-animal approaches, to further reduce uncertainty in the results obtained.

Data source(s) including their availability/accessibility

Agencies can collect this information in dossiers achieved within legal procedures, such as authorization.

Methodology for finding, processing and interpreting the data

Each agency can set up a list of non-animal data collected that often lead to further testing in the animal model, report on the number of animals used per substance assessed for each relevant endpoint or information requirement. As far as possible, there can also be provided information on the reasons the animal testing was considered necessary to perform, e.g. the substance was considered a difficult-to-test substance.

Limitations in the interpretation of the indicator

The best baseline for this indicator would be to understand the use of the animal methods prior to the introduction of the non-animal approach.

Reporting (frequency and timing)

Starting 2027 and to be repeated every year

Example 6

Short title of indicator

Bias in research funding towards the use of animal tests over non-animal approaches

Owner

Research funders

Definition

Number of grants fostering non-animal studies and number grants fostering in vivo studies

Expected utility

To understand when there is bias towards animal methods when funding a research project, to be able to for example provide guidance how to avoid such biases.

Data source(s) including their availability/accessibility

Investigating use of animal methods and non-animal approaches in funded projects, usually publicly available information.

Methodology for finding, processing and interpreting the data

All EU funding of relevant research that can be related to the development of test methods for chemical safety assessment must be listed, and the use of animal versus non-animal testing suggested in funded projects can be collected. It is useful also to collect information on endpoints, type of chemicals and whether non-animal approaches are used to enhance or replace animal testing, which could lead to classify the outcome in three categories: only animal methods, non-animal approaches to enhance animal testing, non-animal approaches to replace animal testing.

Limitations in the interpretation of the indicator

Full understanding for the selection of testing.

Reporting (frequency and timing)

Starting 2027 and to be repeated every year

Example 7

Short title of indicator

Phasing out animal testing carried out fulfilling EU regulatory information requirement

Owner

JRC - European Commission for testing performed with the EU; Industry/Agencies for testing performed outside the EU

Definition

Number of animals used for regulatory testing fulfilling requirements under EU legislation within the 15 legal areas in scope of the roadmap (each legal requirement can be specified and followed as such, why this indicator represents a group of sub-indicators)

Expected utility

Measuring progress towards the phasing out of animal testing and gaining confidence in non-animal approaches within specific legal areas and for specific requirements, that for example can inform other similar requirements in other legal areas, to use a similar approach to the phasing out of animal testing.

Data source(s) including their availability/accessibility

ALURES, publicly available with data from all Member States collected on a yearly basis; Industry/Agencies have information on tested substances

Methodology for finding, processing and interpreting the data

This data is already processed on a yearly basis by the JRC and presented in the European Commission Summary Report on the statistics on the use of animals for scientific purposes in the Member States of the European Union and Norway for each year (example the 2022 summary report: (https://circabc.europa.eu/ui/group/8ee3c69a-bccb-4f22-89ca-277e35de7c63/library/051e5787-7746-46cf-8a0d-310f84fd1900/details?download=true). It might be that Agencies or Industrial associations have some information on testing made outside the EU, or means to collect it.

Limitations in the interpretation of the indicator

Focusing only on statistics collected with the EU, this indicator is not providing any information whether the assumed phased out animal testing, was and is still performed outside the EU, why it is useful to combine this knowledge with available knowledge on testing performed outside the EU.

Reporting (frequency and timing)

Starting 2020 (including 27 Members States and Norway), and to be repeated every year

Example 8

Short title of indicator

Reduction of severity in animal procedures where these tests are still performed

Owner

JRC - European Commission

Definition

Severity scores, horizontal and per legislative sector, and type of test

Expected utility

Measuring progress towards the phasing out of the most severe animal testing by creating awareness, which are the most severe animal procedures used and to which extent in chemical safety assessment.

Data source(s) including their availability/accessibility

ALURES, publicly available with data from all Member States collected on a yearly basis

Methodology for finding, processing and interpreting the data

This data is already processed on a yearly basis by the JRC and presented in the European Commission Summary Report on the statistics on the use of animals for scientific purposes in the Member States of the European Union and Norway for each year (example the 2022 summary report: (https://circabc.europa.eu/ui/group/8ee3c69a-bccb-4f22-89ca-277e35de7c63/library/051e5787-7746-46cf-8a0d-310f84fd1900/details?download=true)

Limitations in the interpretation of the indicator

Focusing only on statistics collected within the EU, this indicator is not providing any information whether severe procedures are phased out within the EU and continued to be performed outside the EU. It could be considered to try to collect data from other sources on the most severe procedures, or the ones decreasing most within the EU.

Reporting (frequency and timing)

Starting 2020 (including 27 Members States and Norway), and to be repeated every year

Annexes

Annex I – Chemical safety assessments under following legislations are in scope of the roadmap

Overview list of regulatory areas and responsible agency

1) Chemicals registered under the REACH Regulation (ECHA)
2) Biocidal Products (ECHA)
3) Plant Protection Products and MRLs for active substances in plant protection products for food and feed (EFSA)
4) Food improvement agents (food additives, food enzymes and food flavourings) (EFSA)
5) Chemicals used in food contact materials (EFSA)
6) Feed additives (EFSA)
7) Human medicinal products (EMA)
8) Veterinary medicinal products and MRLs for active substances in veterinary medicinal products for food-producing animals (EMA)
9) Medical devices
10) Chemicals used in materials/products in contact with drinking water (ECHA)
11) Chemicals covered by the CAD and CMRD (ECHA)
12) Chemicals used in human nutrition (EFSA)
13) Detergents
14) Classification, labelling and packaging of chemicals (ECHA)[118]
15) Water and Waste legislation (identification of priority substances) (ECHA)

[118] The CLP Regulation is based on available data and does not trigger testing. However, it is considered to be in scope of the roadmap since classification relies on data generated under other pieces of legislation, with extensive guidance available that is providing information which data can be used and how. Furthermore, the CLP Regulation is considered to be key for switching to a regulatory hazard assessment based only on non-animal testing because of the classification criteria that it stipulates.

Legislations in more detail and sorted by agency responsibility (with requirements categorisation[119])

ECHA remit:

- Registration, Evaluation, Authorisation and Restriction of Chemicals (**REACH**), Regulation (EC) No 1907/2006 [A]
- Classification, Labelling and Packaging (**CLP**) Regulation (EC) No 1272/2008 [C]
- Biocidal Products Regulation (**BPR**) (EC) No 528/2012 concerning the making available on the market and use of biocidal products [A]
- Drinking Water Directive (**DWD**) (EU) 2020/2184 on the quality of water intended for human consumption [B]
- Water Framework Directive 2000/60/EC [B]
- Environmental Quality Standards Directive 2008/105/EC (EQSD) [B]
- Groundwater Directive (2006/118/EC) [B]
- Occupational exposure limit (OEL) values under Chemical Agents Directive 98/24/EC (**CAD**) and Carcinogens, Mutagens or Reprotoxic Substances Directive 2004/37/EC (**CMRD**) [C]
- The Prior Informed Consent Regulation (EU) 649/2012 (**PIC**), governing the trade of certain hazardous chemicals that are banned or severely restricted in the EU [C?]
- Regulation (EU) No 2019/1021 on persistent organic pollutants (**POPs**) [C?]
- The Waste Framework Directive 2008/98/EC (**WFD**) [B]
- Toys Safety Regulation (in adoption process) (currently Directive 2009/48/EC) [C]

EFSA remit:

- Regulation (EC) No 178/2002 (the General Food Law Regulation), amended by Regulation (EU) 2019/1381 on the transparency and sustainability of the EU risk assessment in the food chain [B]

Feed and feed additives

- Regulation (EC) No 767/2009 on the placing on the market and use of feed [B]
- Feed Additive Regulation (EC) No 1831/2003 [B]
- Regulation (EC) No 429/2008 on the Rules for Implementing the Regulation (EC) No 1831/2003 [A]
- Regulation (EU) 2019/4 on the manufacture, placing on the market and use of medicated feed [C]

Food contact materials

- Regulation (EC) No 1935/2004 on materials and articles intended to come into contact with food [B]
- Commission Regulation (EU) No 10/2011 on plastic materials and articles intended to come into contact with food [B]
- Commission Regulation (EC) No 450/2009 on active and intelligent materials and articles [B]

[119] [A] legislation specifying concrete information requirements and distinct methods
[B] legislation specifying general information requirements, where methods are specified in guidance or selected based on interaction with agency case by case
[C] legislation not specifying specific nor general information requirements but requiring action on the basis of available data

Food improvement agents

- Regulation (EC) No 1331/2008 on authorisation procedure for food additives, food enzymes and food flavourings [C]
- Commission Regulation (EU) No 234/2011 implementing Regulation (EC) No 1331/2008 [B]
- Regulation (EC) No 1332/2008 on food enzymes [B]
- Regulation (EC) No 1333/2008 on food additives [B]
- Regulation (EC) No 1334/2008 on flavourings and certain food ingredients with flavouring properties for use in and on foods [B]
- Regulation (EC) No 2065/2003 on smoke flavourings [B]
- Directive 2009/32/EC on the approximation of the laws of the Member States on extraction solvents used in the production of foodstuffs and food ingredients [B]

Contaminants in food and feed

- Directive (EEC) 315/93 for contaminants in food
- Directive 2002/32/EC for undesirable substances in feed

Nutrition applications

- Regulation (EC) No 1925/2006 on the addition of vitamins and minerals and of certain other substances to foods [C]
- Regulation (EC) No 108/2008 amending Regulation (EC) No 1925/2006 on the addition of vitamins and minerals and of certain other substances to foods [C]
- Directive 2002/46/EC on the approximation of the laws of the Member States relating to food supplements [C]
- Regulation (EU) No 609/2013 on food intended for infants and young children, food for special medical purposes, and total diet replacement for weight control [B]
- Regulation (EU) 2015/2283 on novel foods [B]
- Regulation (EC) 1829/2003 on genetically modified food and feed [B]

Pesticides

- Regulation (EC) No 1107/2009 concerning the placing of plant protection products on the market [B]
- Commission Regulation (EU) No 283/2013 on the data requirements for active substances [A]
- Commission Regulation (EU) No 284/2013 on the data requirements for plant protection products [A]
- Commission Regulation (EU) No 546/2011 on uniform principles for evaluation and authorisation of plant protection products [C]
- Regulation (EC) No 396/2005 on maximum residue levels of pesticides in or on food and feed of plant and animal origin [B]
- Commission Implementing Regulation (EU) No 2020/1740 on the provisions necessary for the implementation of the renewal procedure for active substances [C]

EMA remit:

Authorisation

- Directive 2001/83/EC on the Community Code relating to medicinal products for human use, as amended [B]
- Regulation (EC) No 726/2004 laying down Union procedures for the authorisation and supervision of medicinal products for human use and establishing the European Medicines Agency. [B]
- Regulation (EC) No 2019/6 on veterinary medicinal products [B]
- Regulation (EC) No 470/2009 on the establishment of residue limits of pharmacologically active substances in foodstuffs of animal origin [B]
- Commission Implementing Regulation (EU) 2017/12 on form and content of the applications and requests for the establishment of maximum residue limits in accordance with Regulation (EC) No 470/2009 [B]
- Regulation (EU) No 2018/782 establishing the methodological principles for the risk assessment and risk management recommendations referred to in Regulation (EC) No 470/2009 [A]
- Commission Regulation (EU) 2017/880 laying down rules on the use of a maximum residue limit established for a pharmacologically active substance in a particular foodstuff for another foodstuff derived from the same species and a maximum residue limit established for a pharmacologically active substance in one or more species for other species, in accordance with Regulation (EC) No 470/2009 of the European Parliament and of the Council [C]

Scientific Committee on Consumer Safety (SCCS) remit:

- Cosmetic Product Regulation (EC) No 1223/2009 [B]
- Regulation (EU) 2023/988 on general product safety (non-food consumer products) [C]

Not under any EU-agency

- Detergents Regulation (EU) 648/2004 and 2026/405 [A]

Medical Devices

- Medical Devices Regulation (EC) No 2017/745 (MDR) [B]

Annex II - Ongoing agency activities (ECHA, EFSA and EMA) that support the phasing out of animal testing for chemical safety assessments

This document outlines the work of the European Agencies (EMA, ECHA and EFSA) in relation to 3Rs[120], with a focus on activities that are aligned with the European Commission's ambition to phase out animal testing for chemical safety assessment. It gives some additional context to the legislative frameworks in scope of the roadmap, and the distinct competencies of each agency. Furthermore, it provides information the ongoing work of the individual agencies and their strategic visions, both in relation to the implementation of the roadmap and the integration of 3Rs methods and **alternative to animal testing**=more generally. Finally, it lists critical areas that are of joint interest to all agencies and opportunities for collaborative actions across sectors. Practically, such joint actions could be supported by the Interagency Working Group on Animal-free Approaches which is proposed as part of the roadmap communication.

EMA

Legal Framework

EMA, in cooperation with the European Medicines Regulatory Network (EMRN), provides the scientific support for the authorisation of human and veterinary medicinal products in accordance with Directive 2001/83/EC and Regulation 2019/6, respectively. Medicinal products encompass chemical pharmaceuticals, which are in scope of the roadmap, but also variety of other medicinal products including biologicals, vaccines, gene therapies, advanced therapy medicinal products and novel therapy veterinary medicinal products, which are not. It is important to note that EMAs overall strategy in relation to 3Rs covers all medicinal products, and not just those within the scope of the roadmap.

Medicinal products may be centrally authorised by the Commission, whereby applications are assessed by EMA and the authorisation is valid in all EU Members States. Medicinal products may also be authorised on a decentralised or national basis which is the responsibility of the national competent authorities (NCAs) in the relevant individual members states (i.e. the EMRN). The data to support marketing authorisation is submitted in a common digital format (electronic common technical document eCTD) and includes quality, non-clinical and clinical information. For human medicinal products (HMPs), animal testing is generally required in the non-clinical section as well as to address quality specifications. Animal testing is also required to generate some safety and clinical data for target animal species in the case of veterinary medicinal products (VMPs). The same applies for consumer safety studies. Here, there is no scope for replacement or phasing out of *in vivo* testing, but the principles of reduction and refinement are addressed as part of the assessment. Non-clinical data for both human and veterinary medicinal products is generally presented for pharmacology, pharmacokinetics and toxicology, and includes both animal and non-animal data. Primary pharmacology studies, used to demonstrate efficacy of the product or the proof-of-concept for its therapeutic use are not in scope of the roadmap, which refers to safety assessment. However, application of the 3Rs is also pertinent to animal testing conducted for this purpose and forms part of the assessment.

The availability of structured clinical data, generated in strictly regulated prospective clinical trials, is unique to the pharmaceutical sector. Furthermore, every medicinal product has a specified target population and dosing regimen and therefore an expected level of exposure. This means that every product must be assessed on a case-by-case basis, and the protection goals are variable in each case. Consequently, the goals and outcomes of the risk assessment differ from those in other chemical sectors. For HMPs, a holistic evaluation of the safety, exposure and efficacy of a chemical is conducted and leads to a benefit risk balance. Products can only be authorised on the basis of a positive benefit risk balance. For VMPs, the focus is also on a benefit risk balance, which includes the health benefit for the treated animal against risks arising from its exposure to the product, for

[120] 3Rs = Replace, Reduction and Refinement of animal testing

the human user and/or consumer, who may be exposed during its administration or to its residues via food of animal origin, and for the environment.

Pharmaceutical legislation provides some minimal requirements for the performance of quality and non-clinical tests that should be performed and submitted in the dossier to support a marketing authorisation. However, these requirements should be read in conjunction with relevant European and international guidance documents including internationally harmonised guidelines specific to medicinal products; the ICH[121] guidelines in the case of HMPs, and VICH[122] in the case of VMPs. Medicines regulators like EMA and FDA may also develop their own guidance documents, including guidelines and reflection papers. Many existing guidances refer to the 3Rs principles and allow some flexibility to incorporate alternative methods where scientifically justified, as well as opportunities to reduce and refine animal use. These opportunities are collated in EMA's Reflection papers providing an overview of the current regulatory testing requirements for human and veterinary medicinal products and opportunities for the implementation of the 3Rs. In addition, Scientific Advice can be sought from EMA or NCAs to discuss deviations from guidance including on novel 3Rs approaches.

For HMPs, an assessment of data (which may include animal studies) is also performed at an earlier stage of development to support the authorisation of clinical trials (Regulation 536/2014). Here, the data requirements are different to those for marketing authorisation given the restricted participants and controlled circumstances under which the investigational product is administered. However, the outcome of the evaluation is still based on a benefit risk balance. The evaluation and authorisation of clinical trials is also the remit of NCAs and not of EMA. The same applies to clinical trials for VMPs.

Finally, it should be noted that a significant proportion of applications for marketing authorisation of medicinal products are for generics or informed consent applications (and biosimilars, though not in the scope of the roadmap), where it is generally accepted that no animal testing is required.

EMA activities on 3Rs

EMA is committed to the application of the 3Rs across the lifecycle of medicinal products and to the integration of alternative to animal testing into the medicines regulatory framework. The agency's strategic vision on 3Rs is underscored by a number of policy documents including the Ema Regulatory Science Strategy[123] and the European Medicines Agencies Network Strategy to 2028[124] and it is supported by a standing 3Rs Working Party[125] (3RsWP). The working party which is part of the non-clinical domain publishes its workplan on a 3 yearly basis, identifying annual priorities which include guidance development, stakeholder engagement and communication, pre-authorisation support to innovative 3Rs method developers, oversight of expert groups and training activities. Some key ongoing activities demonstrating alignment with the roadmap objectives are outlined below. Readers are encouraged to consult the linked documents for further details, as well

[121] ICH: The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
[122] VICH: International Cooperation on Harmonisation of Technical Requirements for Registration of Veterinary Medicinal Products
[123] https://eceuropaeu.sharepoint.com/teams/GRP-Interservicegroup-roadmaponnon-animaltesting/Shared Documents/GROW ENV folder/SWD draft/Regulatory Science Strategy to 2025
[124]https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwin1Jna8MmQAxUR2QIHHbqhPfYQFnoECA0QAQ&url=https%3A%2F%2Fwww.ema.europa.eu%2Fen%2Fdocuments%2Fother%2Fseizing-opportunities-changing-medicines-landscape-european-medicines-agencies-network-strategy-2028-final_en.pdf&usg=AOvVaw1o2XJmWqi6QeKzacAzcm09&opi=89978449
[125] https://www.ema.europa.eu/en/committees/working-parties-other-groups/chmp/3rs-working-party

as the EMA 3RsWP[126], ethical use of animals in medicines testing[127] and regulatory acceptance of NAMs[128] webpages.

Guidance activities:

EMA systematically implements the 3Rs concept and refers to the use of alternative to animal testing in all of its relevant guidance documents. The main goals of ongoing guidance activities are to provide an inventory of regulatory testing requirements involving animal use and describe opportunities for 3Rs implementation, and to provide a framework for acceptance of novel 3Rs approaches including a gradual phase in of alternative to animal testing.

- Cyclic revisions of the "Reflection papers providing an overview of the current regulatory testing requirements for medicinal products for human[129] or veterinary[130] use and opportunities for implementation of the 3Rs"
- Revision of the "Guideline on the principles of regulatory acceptance of 3Rs (replacement, reduction, refinement) testing approaches[131]" as outlined in the Concept Paper[132]

Stakeholder engagement:

Core 3RsWP stakeholders include pharmaceutical industry associations, NGOs, EU and international research consortia, ECHA, EFSA and OECD. Observers include EC services (DG SANTE, DG RTD, DG ENV, DG GROW, DG JRC (EURL ECVAM)) and EDQM. In recognition of the global context of medicines development and the need for international harmonisation on 3Rs recommendations, EMA led in the establishment of an International Medicines Regulators' Working Group on 3Rs (IMRWG3Rs).

- Annual Interested Parties meetings (reports available for 2023[133], 2024[134] and 2025[135])
- International Medicines Regulators' Working Group on 3Rs (terms of reference available[136])

Regulatory acceptance of NAMs – supports and procedures (safe spaces model, overview available[137]):

EMA fosters regulatory acceptance of alternative to animal testing by offering various mechanisms by which method developers can interact with EMA at different stages of development.

- Innovation Task Force briefing meetings
- Qualification of novel methodologies
- Scientific advice
- Portfolio and technology meetings

[126] https://www.ema.europa.eu/en/committees/working-parties-other-groups/chmp/3rs-working-party
[127] https://www.ema.europa.eu/en/human-regulatory-overview/research-development/ethical-use-animals-medicine-testing
[128] https://www.ema.europa.eu/en/human-regulatory-overview/research-development/ethical-use-animals-medicine-testing/regulatory-acceptance-new-approach-methodologies-nams-reduce-animal-use-testing
[129] https://www.ema.europa.eu/en/overview-current-regulatory-testing-requirements-medicinal-products-human-use-opportunities-implementation-3rs-scientific-guideline
[130] https://www.ema.europa.eu/en/overview-current-regulatory-testing-requirements-veterinary-medicinal-products-opportunities-implementation-3rs-scientific-guideline
[131] https://www.ema.europa.eu/en/regulatory-acceptance-3r-replacement-reduction-refinement-testing-approaches-scientific-guideline
[132] https://www.ema.europa.eu/en/documents/scientific-guideline/concept-paper-revision-guideline-principles-regulatory-acceptance-3rs-replacement-reduction-refinement-testing-approaches_en.pdf
[133] https://www.ema.europa.eu/en/documents/report/report-3rswp-public-session_en.pdf
[134] https://www.ema.europa.eu/en/documents/report/report-3rswp-public-session-2024_en.pdf
[135] https://www.ema.europa.eu/en/documents/report/report-3rs-working-party-3rswp-plenary-meeting-april-2025_en.pdf
[136] https://www.ema.europa.eu/en/documents/other/terms-reference-tor-international-medicines-regulators-working-group-3rs_en.pdf
[137] https://www.ema.europa.eu/en/human-regulatory-overview/research-development/ethical-use-animals-medicine-testing/regulatory-acceptance-new-approach-methodologies-nams-reduce-animal-use-testing

- Voluntary submission of data ("safe harbour") procedure

Oversight of expert groups and training activities:

The 3RsWP oversees the work of an operational expert group on the application of the 3Rs in batch release and quality control of medicines and European Specialised Expert Community on non-clinical and NAMs which provides a platform for interaction between academic and regulatory experts. In addition, EMA provides training to regulators within the EMRN on 3Rs topics relevant to regulatory assessment of medicines via the EU-Network training centre.

- Batch release operational expert group
- Non-clinical and new approach methodologies ESEC[138]
- EU Network Training Centre[139]

ECHA

Legislative framework

The primary objective of EU legislation regulating industrial chemicals is to provide a high level of protection of human health and the environment. It is underpinned by the identification of hazardous properties of substances, with REACH and CLP as the essential horizontal EU Regulations.

Since its entry into force in 2007, REACH is the regulatory framework producing the largest knowledge base on chemicals globally. REACH ensures that industry provides adequate data, using as a last resort tests on animals, to assess the hazardous properties of chemicals. CLP enables the classification of chemicals based on the adverse effects observed using standard testing methods, including tests on animals, and provides a framework for generic risk management. This system provides predictability and legal certainty for industry compliance and a framework for authorities to enforce.

In addition to REACH and CLP, sectorial legislation complements the EU framework on the identification of hazards, risk assessment and management of chemicals. Chemicals management relies on the effective interplay between REACH, CLP and the sectorial legislation. It depends on the cooperation of different actors (industry, authorities, third parties, etc.) according to the roles and responsibilities described by the regulatory frameworks, providing predictability and legal certainty for industry to comply and authorities to enforce.

The generic system under REACH and CLP does not require the intervention of authorities by default – chemicals are not systematically assessed for safety by the regulators before they are put on the market; this assessment is done solely by industry ("the reversal of the burden of proof"). However, authorities may scrutinise these assessments and take regulatory actions (e.g. request further information or initiate regulatory risk management), where required. In addition, this system allows adaptations to standard information requirements to be used for supporting hazards and risk assessment.

ECHA activities on alternatives to animal testing

ECHA is committed to promoting alternatives to animal testing and, within its mandate, has been actively addressing this topic since the agency's establishment.

The current work can is structured around the following main areas:

[138] https://www.ema.europa.eu/en/committees/working-parties-other-groups/chmp/non-clinical-new-approach-methodologies-european-specialised-expert-community
[139] https://www.ema.europa.eu/en/about-us/how-we-work/european-medicines-regulatory-network/eu-network-training-centre-eu-ntc

1. Regulatory activities and use of alternative methods in ECHA's processes
2. Translation of NAMs into regulatory applications
3. Input to harmonisation and reporting
4. Computational methods and data
5. Trainings and promotion of proper use in the regulatory context
6. Interactions with key stakeholders

Regulatory activities and use of alternative methods in ECHA's processes

This work area reflects the current integration of New Approach Methodologies (NAMs) into ECHA internal regulatory processes, such as evaluation, and outlines further implementation to support initiatives like chemical grouping. It also addresses legal obligations, including reporting on alternatives to animal testing.

Key activities include:

- Assessment of QSAR predictions in ECHA processes, primarily through expert consultations for REACH and Biocides
- Publication of the Article 117(3) report on the use of alternatives to animal testing under REACH
- Support for drafting and revising ECHA guidance involving alternatives to animal testing

Translation of NAMs into regulatory applications

This work area aims at providing regulatory input to flagship (research) projects aiming at developing alternatives to animal testing suitable for our regulatory needs. It also enables joint efforts with authorities from Europe and other regions such as the US or Canada to work together towards the identification and acceptance of alternatives to animal testing in regulatory frameworks

Key activities include:

- Implementation of Framework contracts for the development of NAM based tools and data for hazard identification and characterisation of industrial chemicals with an emphasis on omics data generation and methodological developments of NAM related to hazard assessment of industrial chemicals (including both toxicodynamic and toxicokinetic assessment)
- Participation in scientific projects which address key aspects for regulatory acceptance. Example collaborations include APCRA (Accelerating the Pace of Chemical Risk Assessment), PARC (Partnership for the Assessment of Risks from Chemicals), and EU research consortium such as ASPIS
- Definition (and communication) of the critical needs to progress towards an animal-free regulatory system for industrial chemicals and expectations for new approach methods (NAMs) – e.g. within the ECHA KARC document (Key areas of regulatory needs)

Input to harmonisation and reporting

This work area is related to the OECD activities important for the further development and regulatory implementation of alternatives to animal testing.

Key activities include:

- Collaboration with WPHA (Working Party on Hazard Assessment) to develop a Guidance on best practices and standardisation of omics samples and the WNT (Working Party for the Test Guideline Program) to include optional omics sampling in selected OECD Test Guideline studies to enable the analysis of additional molecular/mechanistic parameters.
- Active participation in OECD Expert groups such as the Advisory group on Emerging Science in chemicals assessment (ESCA), the 'omics Expert Group and the QSAR Expert Group.
- Development of the OECD QSAR Assessment Framework (QAF) and example case

studies.

Computational methods and data

This work area focuses on making ECHA registration data available for developers and scientists.

Key activities include:

- Continuous development and maintenance of the OECD QSAR Toolbox.
- Promotion of the OECD QSAR Assessment Framework within academia and industry to enhance regulatory acceptance of QSAR predictions.
- Provision of toxicological data in the harmonised format IUCLID for further development of alternatives to animal testing.
- Provision of curated datasets with experimental data for selected endpoints, e.g. acute toxicity; Adult neural toxicity, bioaccumulation and aquatic toxicity.
- Assessing the predictive capacity of environmental QSARs and predictive models to support prioritisation and scientific decision making further implemented.
- Assessment of QSARs reliability for selected regulatory endpoints.
- Horizon scanning of mature (computational) methods, technologies and tools of relevance to ECHA.

Organisation of trainings and promotion of proper use of NAMs in the regulatory context

This work area is aiming at sharing ECHA experience and knowledge with ECHA staff, stakeholders and registrants.

Key activities include:

- Trainings on NAMs for ECHA staff, Member States, committees, industry, and scientific community
- Public webinars on application of QSARs, QSAR assessment framework and PBK modelling
- Answering speaking requests and presentation of ECHA activities in dedicated fora (e.g. conferences, working groups)

Interactions with stakeholders

This work area is aiming at exchange of knowledge, experience and alignment in the areas of the key importance for the European institutions. In addition, dialogue with NGOs is also covered here.

Key activities include:

- Contribution to the roadmap on phasing out animal testing for chemical safety assessments.
- Collaboration with Member States on activities related to alternatives to animal testing.
- Alignment with sister agencies, especially EFSA and EMA, and joint projects on guidance development
- Representing ECHA at the European Partnership for Alternative Approaches to Animal Testing (EPAA).
- Dialogue with NGOs – presentation of plans and reply to questions.
- Dialogue with industrial partners and CROs involved in routine alternatives to animal testing data generation.

EFSA

Legislative framework

The European Food Safety Authority (EFSA) has been established under the General Food Law Regulation (EC) No 178/2022[140] that lays down the procedures in matters of food safety. The main aim of the regulation is to ensure *high level of protection of human life, consumers' interests and environment in relation to food, while ensuring the effective functioning of the internal market*.

EFSA provides independent scientific advice on existing and emerging risks in the food chain to help protection of consumer and environment covering the remits of i) food and feed safety, ii) nutrition, iii) animal health and welfare, iv) plant protection and v) plant health.[141] EFSA follows specific regulations under the different legal frameworks covered by its activities. EFSA's Scientific Committee[142] has the task of supporting the work of the different Scientific Panels on cross-cutting scientific issues, focusing on the development of harmonised risk assessment methodologies. EFSA's Scientific Panels[143] are responsible for providing the scientific opinions of the Authority, each within their own spheres of competence. It is also within EFSA's remit the independent and timely risk communication regarding food and feed safety as well as the promotion of scientific cooperation.

Since March 2021, EFSA complies with Regulation (EU) 2019/1381 on transparency and sustainability of the EU risk assessment in the food chain, keeping the highest standards of transparency. Overall, EFSA publishes around 500 outputs on a yearly basis covering generic mandates as well as applications for regulated products, which are governed by more than 34 different EU Directives and Regulations and follows 39 different workflows. These outputs relate to include plant protection products, plant health, food and feed additives, animal health and welfare, food flavourings, biological hazards, feed additives, food additives and flavourings, food enzymes, food contact materials, nutrition and novel food innovation foods, genetically modified organisms, and contaminants in the food and feed chain[144].

EFSA's activities on alternatives to animal testing

EFSA is committed to implement the principles of the 3R's and integrating alternatives to animal testing in food and feed risk assessment, by working closely with the alternatives to animal testing community and exploring existing technology and legislative possibilities[145] (EFSA SC, 2009[146]; EFSA, 2022a[147]; EFSA, 2022b[148]; EFSA, 2023[149]; EFSA, 2025[150]).

EFSA has as a strategic objective, the development and integration of alternatives to animal testing and omics in regulatory risk assessment as a means to ensure preparedness for future risk analysis (EFSA Strategy 2027, 2021[151]). The possibility to use alternatives to animal testing is cited in a number of EFSA's guidance documents including novel foods, food additives or the general applications of read-across in chemical risk assessment (EFSA NDA, 2024[152]; EFSA guidance on food additives; OECD, 2023, EFSA SC, 2025[153]). Furthermore, EFSA has worked with the alternatives to animal testing community to develop roadmaps as to increase the incorporation of NAMs in food and feed risk assessment, making their use when relevant and possible to complement the risk assessment). In some regulatory areas, within EFSA's remit (such Pesticides; Feed and feed additives or Food contact material) there are established legal requirement that list

[140] https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32002R0178
[141] https://www.efsa.europa.eu/en/topics
[142] https://www.efsa.europa.eu/en/science/scientific-committee-and-panels/scientific-committee
[143] https://www.efsa.europa.eu/en/science/scientific-committee-and-panels
[144]More details are available in https://www.efsa.europa.eu/en/topics
[145] More information is available in https://www.efsa.europa.eu/en/topics/topic/alternatives-animal-testing and https://www.efsa.europa.eu/en/topics/topic/methodology.
[146] https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2009.1052
[147] https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2022.e200502
[148] https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/sp.efsa.2022.e200506
[149] https://www.sciencedirect.com/science/article/pii/S0924224423000572
[150] https://efsa.onlinelibrary.wiley.com/doi/full/10.2903/j.efsa.2025.e220401
[151] https://www.efsa.europa.eu/sites/default/files/2021-07/efsa-strategy-2027.pdfhttps:/www.efsa.europa.eu/sites/default/files/2021-07/efsa-strategy-2027.pdf
[152] https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2024.8961
[153] https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2025.9586

the type of studies to be used in the assessment. In these cases, moving to different type of requirements, including the move to a complete animal-free assessment relying more on alternatives to animal testingmay require additional work for both jurisdictions (adjusting legal frameworks) as well as risk assessors and academia (adjusting guidance documents and validating methodologies).

The relevance of alternatives to animal testingin EFSA's work is further highlighted in the EFSA 2027 strategy[154] o "*develop and integrate new scientific developments focusing on NAM-based methods and minimisation of animal testing, innovations in food systems, data, and technology, and strive to meet One health policy needs*". This has triggered the development of a series of NAMs' activities in EFSA that can be grouped into the following major areas:

- **Closing gaps and support scientific discovery on NAMs**

Regarding the challenge of closing gaps between science and regulation, EFSA has up to now invested over € 30 million euros in projects related to the use of NAMs in food and feed risk assessment. In 2022, EFSA established the Innovative *Risk Assessment* Methodologies programme (IRMA) and the Data. & Evidence programme (DEEP) s to ensure preparedness and knowledge of up-to-date risk assessment methodological developments. These programmes also aim to meet new scientific challenges that EFSA is called to face in the coming five years in food, feed, and environmental safety assessment[155].

EFSA's funded project address NAMs' tools for projects in topics such as inter-human variability in toxicodynamics, kinetics and metabolism, neurodevelopment, endocrine disruptors, environmental risk assessment, pesticide metabolites, nanomaterials in food and feed, protein safety, toxicogenomics and transcriptomics, artificial intelligence and in silico model development including quantitative structure–activity relationship (QSAR) and physiologically-based kinetic models (PBK). These projects target unique needs of regulatory science within EFSA's remit. They require ongoing updates and validation that connect theory with practice.

- **Supporting NAMs inclusion and assessment in EFSA's work**

A number of working groups (WGs) have been and/or are being established at EFSA to develop alternatives to animal testing strategies such as benchmark dose modelling, read-across, developmental neurotoxicity in vitro testing battery, use of quantitative in vitro to *in vivo* extrapolation (QIVIVE)/PBK modelling, phasing out dog studies, use of TK-TD models in environmental risk assessment, ADME testing and PBK modelling for human and environmental risk assessment. At the same time, EFSA has established several WGs that give support to different process and projects where the use of alternatives to animal testing is discussed such as the WG on Endocrine Disruptors, genotoxicity, Neurotoxicity and the Scientific Committee working group on NAMs.

The implementation of the One Substance, One Assessment (OSOA) approach and the development of an EU-based data platform will significantly support the reduction of animal testing. These initiatives support data sharing and harmonisation, preventing duplicate studies on the same substance being conducted under different regulatory remits.

EFSA collaborates and exchanges data, expertise and methodologies with sister agencies and relevant stakeholders such as the Organisation for Economic Co-operation and Development (OECD), in order to reach consensus on the applicable criteria/standards for alternatives to animal testing- data and their integration for use in feed and food risk assessment. For example, EFSA's OpenFoodTox 3.0 project delivered a workflow within IUCLID that will make EFSA and ECHA data available under the same format, allowing stakeholders to consult such data and have a better understanding on how studies were used in the risk assessment. Bioaccumulation models have also been developed over the years and support assessment of persistence of chemicals in humans and terrestrial and aquatic compartments such as bioconcentration factors (BCF) and bioaccumulation factors (BAF). EFSA has also contributed to the development of an OECD Harmonised Template

[154] https://www.efsa.europa.eu/en/corporate-pubs/efsa-strategy-2027-science-safe-food-sustainability
[155] For more details, please see https://www.efsa.europa.eu/en/topics/topic/methodology

structure as to facilitate the reporting and structuring of QSAR data in the IUCLID database used by ECHA (OECD GD 386, 2023[156]). The aim of these activities is to improve the availability of standardised data within a harmonised framework for regulatory assessment. Finally, EFSA has actively contributed to the OECD Initial recommendations on the Evaluation of Data from Developmental Neurotoxicity (DNT) In-vitro Testing Battery (OECD 2023[157]).

- **EFSA's Knowledge and Innovation Community (KIC) on NAMs**

The NAMs Knowledge and Innovation Community (KIC) aims to create a common culture to foster cross-area use of NAMs within EFSA One of the objectives of the KIC on NAMs is to understand and share knowledge across EFSA community and beyond on the use of NAMs within EFSA's regulatory and other regulatory frameworks, identify gaps in the EFSA strategy on NAMs and how these NAMs are used/approached by the Units in their day-to-day risk assessments. Activities include contribution to define and prioritise future EFSA projects on NAMs, being the vehicle to articulate collaboration also with the NAMs ecosystem outside EFSA including the implementation of the roadmap of the European Commission.

- **Strengthening NAMs cooperation beyond EFSA**

EFSA is involved in many OECD activities in relation to NAMs particularly with the Working Party on Hazard Assessment. These activities include supporting OECD working groups including harmonised templates, e-chem portal, QSAR reporting framework (QSAR RF), developmental neurotoxicity in vitro batteries, AOP development and update of the guidance document of the use of PBK models for regulatory purposes. EFSA is also collaborating with the EPAA as members of the Steering Committee as well as the international consortiums ILMERAC (International Liaison Group on Methods for Risk Assessment of Chemicals in Food), Accelerating the Pace of Chemical Risk Assessment (ACPRA). In addition, over the last few years, EFSA has had active collaboration with the US-EPA and FDA particularly in the area of in silico modelling. In June 2025, EFSA has also participated to the Singapore WHO workshop on NAMs held from which a chapter on NAMs will be introduced in the Environmental Health Criteria 240 describing the principles of chemical risk assessment.

Finally, EFSA is very active in the NAM area within the PARC and ASPIS projects reviewing PARC projects and outputs and stimulating the use of PARC data and in silico models for EFSA 's ongoing risk assessments (e.g. mycotoxins, pesticides, contaminants, phthalates etc).

- **Training and Guidance**

EFSA organises regular training courses on NAMs to EFSA staff and to EFSA's experts from its Scientific Panels and Scientific Committees (i.e. NAMs, TKPlate, genotoxicity, QSAR, BMD, ED). EFSA also provides training for other international agencies (US-EPA, FDA, NIEHS), to Member States, Article 36 organisations and to research consortiums such as PARC. In addition, EFSA is developing several guidances to support applicants, Member States and the food and feed community on how to include alternatives to animal testingin risk assessment. This includes tiered approaches in area specific guidances, but also cross-area guidances such as weight of evidence and read-across.

[156] https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/11/q-sar-assessment-framework-guidance-for-the-regulatory-assessment-of-quantitative-structure-activity-relationship-models-and-predictions_9b064821/d96118f6-en.pdf
[157] https://www.oecd.org/en/publications/initial-recommendations-on-evaluation-of-data-from-the-developmental-neurotoxicity-dnt-in-vitro-testing-battery_91964ef3-en.html

Critical areas and joint actions agreed between ECHA, EFSA and EMA

Several key areas have been identified as being crucial for the transition to non-animal approaches and of common interest to EFSA, EMA, and ECHA. These include but are not limited to:

- Development of reliable ADME (absorption, distribution, metabolism and elimination) profiling and PBK (physiologically based kinetic) modelling tailored to regulatory needs.
- Establishing reliable (Q)IVIVE ((quantitative) in vitro to *in vivo* extrapolation) and robust dose-response modelling.
- Enhancing the use of in silico methods (primarily (Q)SAR ((quantitative) structure-activity relationships) modelling) with more explicit criteria for the domain of applicability.
- Integration of novel methods, approaches or combinations of methods in (reproductive) toxicology assessment.
- Reduction of the number of species tested, where justified.

Across these areas, the Agencies are committed to identifying commonalities, learning from each other as to method applicability in the different areas of work, and working together to ensure efficient and safe uptake of the models as appropriate in each of the different sectors. Each Agency will identify existing support, and new opportunities, for these developments within their sector.

Example Actions

Short-Term Replacements:

- Use of in vitro kinetic data and PBK models to refine testing strategies in chemical risk assessments, reducing or limiting the dependence on chronic and sub-chronic animal tests (while noting that non-clinical PK also informs clinical trials in the pharmaceutical sector, which may result in differing requirements).
- Supporting the implementation of the OSOA approach, keeping up to date the Openfoodtox database and the EU data platform to facilitate data sharing thus preventing the conduct of redundant studies, particularly when the same substances are being assessed under different frameworks.
- The use of bioaccumulation models and QSAR models.
- Establishing validation/qualification frameworks to standardise the evaluation of alternatives to animal testing, supporting global compliance/ trade harmonisation.

Long-Term Developments/Replacements:

- Developing a substitution plan for invertebrate studies in ERA.
- Transitioning methodologies to focus on AOPs and KE rather than traditional animal endpoints, with the goal of combining multiple methods to address regulatory concerns.

Annex III - Engaging with the general public

The importance of Public Engagement and More Inclusive Stakeholder Representation (perspective of the SAFE Consortium[158])

Rationale and Background

The European Commission's roadmap, inspired by the European Citizen Initiative (ECI) *Save cruelty Free Cosmetics, and calls from the European Parliament to accelerate the transition* gives a tremendous opportunity to strengthen the democratic spirit which underpins European values and institutions. Engagement with European citizens is a strength for the roadmap implementation process. By involving citizens, including patient and consumer organisations, directly in the policy process a number of direct and indirect benefits can be gained.

Effective implementation

As a democratic institution, the EU fundamentally serves the interests of its citizens. Instead of solely relying on opinion polls, engagement with the general public will make sure that policy is more tailored to their desires and needs and will ensure that the resources spent towards developing and implementing the strategy of the roadmap will be implemented as efficiently and effectively as possible.

Interests in society are diverse and cannot be easily generalized. Engagement is an effective way to shorten the transmission of ideas and increase the representation of various groups and their viewpoints whilst ensuring that the values of the people being represented are reflected in policy implementation. Ensuring broad participation can ensure that there is not undue influence from special interest groups, such as commercial interests. Broad participation also provides a means of representing the interests of non-human animals (Donaldson, 2020), see section 3 below.

Strengthening EU institutions

Consultation with citizens can inspire a sense of participation in the implementation process of the roadmap, but it also indirectly strengthens trust in EU institutions themselves. This helps address what has been referred to as a "democratic deficit" in EU institutions.

Strengthening trust in science and evidence-informed policy

Engagement will also help empower citizens in the science-policy process to become advocates against disinformation and misinformation and increase trust in the science-policy process (Grant, 2023; Janara and Donaldson, 2023; McGlacken and Hobson-West, 2022).

Producing better science

Opening the policy process to public engagement and consultation can help guide the direction of science inquiry which could help overcome possible biases in the existing scientific paradigm based on animal models. Involving citizens can help scientists formulate research questions which align with the needs and concerns of citizens. For example, in the healthcare domain, the James Lind Alliance (JLA) Priority Setting Partnerships (PSPs)[159] are collaborative projects that bring together patients, carers, and clinicians to identify and prioritise the most important unanswered questions.

Science based on the collaboration and engagement of citizens can also be used to facilitate data collection in large-scale studies. For example, it could indirectly help human biomonitoring across the EU, enabling more comprehensive and representative datasets to drive evidence-based decision-making.

[158] Accelerating the transition to animal-free NGRA: A transformative governance approach | NWO
[159] https://www.jla.nihr.ac.uk/about-priority-setting-partnerships

Inspiring Innovation

Maintaining and strengthening public awareness of chemical safety assessment not only fosters a deeper understanding of its importance but also helps inspire and attract the next generation of scientists. By engaging communities in this critically important field, we can spark curiosity, drive innovation, and cultivate a diverse talent pool ready to tackle future challenges.

Forms of Public Engagement

Multiple strategies exist to effectively engage with the general public (ESIC, 2025). Citizen assemblies (Patberg, 2024) and Citizen Juries (Setälä et al, 2023) are methods which see engagement with multiple people in a deliberative forum in order to openly discuss policy questions. This kind of engagement has been shown to enhance trust in institutions, despite polarizing opinions. It has for example been used in relation to the Conference on the future of Europe, and a report is available on how a European Citizen's Assembly could be set up (Abels et al. 2022). Special care needs to be given in how these forms of public engagement are constructed to ensure that the full diversity of citizens is represented. The inclusion of citizens who reflect the diversity of a given country or region as well as patient and consumer organisations should be considered.

In addition to these participatory methods, more classic information campaigns can be used to engage citizens in safety assessment. Citizens are often not aware of the wide range of applications of animal testing for chemical safety. In order for citizens to be able to make informed decisions and play a productive role in policy processes, a basic understanding of the policy and its implications is needed (Holmberg and Ideland, 2012). The EC and EU agencies could for example work together with civil society organizations to implement information campaigns.

Also, the transition to animal-free safety assessment is the only transition (as compared to e.g., the energy or food transition) in which the consumer has no apparent role. One way could be by providing consumers information on which products have been developed using animal testing (e.g. pesticide residues in food products), through labelling. This market-based approach can support public engagement in the transition by providing animal-free options to European consumers. In this manner, consumers can support the acceleration of the transition by choosing animal-testing free options.

Representation of non-human animals

Many (future) non-human animals obviously have high stakes in the transition towards animal free safety assessment and should therefore, from a democratic perspective, be included in decision-making processes. While it is difficult to *directly* include non-human animals within current democratic processes there are nevertheless ways to include non-human animal interests through *indirect* representation and through combinations of direct and indirect representation.

Forms of Non-human Animal Representation

Indirect representation by human proxies in citizens' assemblies/juries or in expert groups can be to the indirect representation of future generations in questions regarding e.g. climate change (Hara et al, 2019). This means regular human citizens are tasked to represent the perspective and interests of a specific group of non-human animals, or an individual animal, rather than their own (Magaña, 2022; van Veen and Helvoirt, 2024). They are given the time to understand the perspective and interests of this group or individual, including through meeting individuals of the group they represent. Thereafter they participate at meetings in the same way as the other representatives (see Van Veen & Van Helvoirt, 2024).

Another form of representation is through **microboards**, whereby a group of people that is close to an individual (human or non-human) forms a small group of people which is then involved in decision making processes (e.g. stakeholder consultations, deliberative processes) relating to a particular issue or task. In the case of animals, this could for example be human caretakers of an animal living in a sanctuary or in a human's home (e.g. a mouse that lives in a sanctuary for mice, which formerly used to live in a laboratory, could be represented by human caretakers working in that sanctuary).

Expert indirect representation is also an alternative. This means appointing humans who are experts in animal welfare, biology or communication, to advocate on their behalf. These experts act as proxies, ensuring that animal interests are considered in decisions that affect them. Examples include: a) **an Ombudsman for Animals**, who investigates complaints and proposes reforms, ensuring that animals are treated ethically in society; b) a **Seat in Parliament for Animals** held by an expert who speaks exclusively for animal concerns, without representing a political party or constituency; c) a **Labour Union for Animals** (De Rooij, 2023) that would campaign for better working conditions for animals used in biomedical research and regulatory testing.

The development of governance models for including the interests of non-human animals in decision making is a promising avenue of research. An example at the organisational level is the Zoöp Nieuwe Instituut in the Netherlands, which could be an inspiration for decision making in other organisations that carry out regulatory testing.

Conclusions

The European Commission's roadmap presents a vital opportunity to enhance democracy, trust in science, policy effectiveness and efficiency by actively engaging citizens and by considering the representation of non-human animals. Through inclusive methods like citizens' assemblies and innovative representation models, the EU can foster better science, inspire innovation, and strengthen public trust in its institutions. Embracing these approaches ensures policies reflect diverse needs while advancing ethical and evidence-based decision-making.

References

Abels, G., Crum, B., Alemanno, A., Demidov, A., Hierlemann, D., Renkamp, A., & Trechsel, A. (2022). *Next level citizen participation in the EU: Institutionalising European Citizens' Assemblies*. Bertelsmann Stiftung. https://research.vu.nl/ws/portalfiles/portal/158204609/Next_Level_Citizens_Participation_in_the_EU.pdf

De Rooij, M. (2023). *Hoeveel vakantiedagen heeft een varken?* Noordboek. https://noordboek.nl/boek/hoeveel-vakantiedagen-heeft-een-varken/

Donaldson, S. (2020). Animal Agora: Animal Citizens and the Democratic Challenge. *Social Theory and Practice*, *46*(4), 709–735. https://doi.org/10.5840/soctheorpract202061296

ESIC (Evidence Synthesis Infrastructure Collaborative). 2025. 'ESIC_citizen Call to Action_30 July 2025'. https://evidencesynthesis.atlassian.net/wiki/spaces/ESE/pages/222167041/Citizen+call+to+action

Grant, W. J. (2023). The Knowledge Deficit Model and Science Communication. In *Oxford Research Encyclopedia of Communication*. https://doi.org/10.1093/acrefore/9780190228613.013.1396

Hara, K., Yoshioka, R., Kuroda, M., Kurimoto, S., & Saijo, T. (2019). Reconciling intergenerational conflicts with imaginary future generations: Evidence from a participatory deliberation practice in a municipality in Japan. *Sustainability Science*, *14*(6), 1605–1619. https://doi.org/10.1007/s11625-019-00684-x

Holmberg, T., & Ideland, M. (2012). Secrets and lies: "selective openness" in the apparatus of animal experimentation. *Public Understanding of Science (Bristol, England)*, *21*(3), 354–368.

Janara, L., & Donaldson, S. (2023). Animal Ghosts at Canadian Universities: The Politics of Concealment and Transparency. *Animals*, *13*(24), 3760. https://doi.org/10.3390/ani13243760

Magaña, P. (2022). The political representation of nonhuman animals. *Social Theory and Practice*, *48*(4), 665–690.

McGlacken, R., & Hobson-West, P. (2022). Critiquing imaginaries of 'the public' in UK dialogue around animal research: Insights from the Mass Observation Project. *Studies in History and Philosophy of Science*, *91*, 280–287. https://doi.org/10.1016/j.shpsa.2021.12.009

Patberg, M. (2024). Farewell to constituent power? The Conference on the Future of Europe, citizens' assemblies and the democratic minimum. *Global Constitutionalism*, *13*(2), 411–428. https://doi.org/10.1017/S2045381723000333

Setälä, M., Serup Christensen, H., Leino, M., & Strandberg, K. (2023). Beyond polarization and selective trust: A Citizens' Jury as a trusted source of information. *Politics*, *43*(4), 472–488. https://doi.org/10.1177/02633957211024474

van Veen, A., & van Helvoirt, B. (2024). *De-Mooo-cracy! A performative experiment with a cow-human citizen's assembly: Evaluation report* (p. 17). http://dx.doi.org/10.13140/RG.2.2.35140.44168

Annex V Summaries of activities of the UK NC3Rs relevant to actions in this roadmap

A) UK NC3Rs report - Refining and replacing acute toxicity tests in chemical safety assessments

Introduction

This report outlines the work of the UK's National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) to avoid, reduce and refine the use of animals in acute toxicity testing for chemical safety assessment. It complements considerations in chapter 4.1. of this SWD.

The UK's NC3Rs is participating in the EPAA Acute Toxicity Taskforce, and in this function helping to shape international policy to promote the adoption of more refined approaches and support the broader integration of non-animal approaches into regulatory frameworks. By enabling scientifically justified reductions in animal use, NC3Rs contribute to the European Commission's ambition to phase out animal testing and advances global efforts to modernise chemical safety assessment while maintaining high standards of human health protection. Additional collaboration with the UK's efforts to phase out animal testing is occurring at he level of the OECD. In November 2025, the UK published their national roadmap to phasing out animal testing, which has a very similar scope as the Commission roadmap.

Background

Acute toxicity studies are among the most severe animal tests still used in regulatory safety assessments. Typically conducted in rats, these studies aim to determine the dose or concentration of a substance that causes death in 50% of the test animals (LD50/LC50). Lethality is the intended endpoint, and the procedures are associated with high levels of animal suffering.

While acute toxicity testing is no longer required for pharmaceuticals – following changes to ICH M3(R2) in 2009 as a result of previous NC3Rs work in this area (Chapman et al., 2010; Robinson & Chapman, 2009; Robinson et al., 2008) – it remains mandatory for chemicals and agrochemicals under various regulatory frameworks. Several OECD test guidelines continue to support these tests globally, including for hazard identification and classification and labelling (CLP).

Validated non-animal approaches exist for predicting acute effects such as skin and eye irritation and sensitisation but replacing oral and inhalation acute toxicity tests remains more complex. While international efforts are underway to develop non-animal approaches, including *in silico* models for oral toxicity and *in vitro* systems for inhalation, these methods have not yet been accepted as stand-alone replacements in regulatory frameworks. Until full replacement is possible, the NC3Rs is leading efforts to refine these tests to reduce animal suffering. This includes promoting the use of "evident toxicity" –

clinical signs that predict lethality – instead of death as an endpoint and influencing international guideline revisions.

NC3Rs Acute Toxicity Project: Progress and Impact

The NC3Rs has played a central role in embedding the 3Rs into acute toxicity testing since its establishment in 2004. Early work focused on removing the requirement for single-dose acute toxicity studies in pharmaceutical development, contributing to changes in ICH M3(R2) (Chapman et al., 2010; Robinson & Chapman, 2009; Robinson et al., 2008). More recently, efforts have shifted to refining acute toxicity tests required for chemicals and agrochemicals, where they are still required.

Currently, there are three OECD test guidelines for acute oral toxicity (TG 420, 423, and 425) and three for acute inhalation toxicity (TG 403, 433, and 436). Of these, only TG 420 and TG 433 allow the use of evident toxicity as an endpoint. The others still rely on lethality, despite the availability of more humane alternatives.

For inhalation studies, the NC3Rs led a large data-sharing initiative and developed guidance that contributed to the adoption of TG 433 in 2017 (Sewell et al., 2015; 2018). This guideline uses evident toxicity and represents a significant refinement and reduced animal numbers. Work continues with partners including the EPAA Acute Toxicity Taskforce and the UK's OECD national coordinator to remove TG 403 (the LC50 method), following the precedent set by the deletion of oral and dermal equivalents (deleted TG 401 and revised TG 402).

In 2024, OECD member countries accepted the NC3Rs proposal to revise TG 403 so that it is only used as a last resort. However, some countries continue to justify its use for determining acute exposure guideline levels. To address this, work is underway to demonstrate that TG 433 can also meet these needs.

For oral studies, TG 420 has permitted the use of evident toxicity since 1992, but uptake remains limited. This is due to a lack of clear guidance and the continued availability of lethality-based alternatives. In collaboration with the EPAA, the NC3Rs published new supporting data in 2024 to validate the use of evident toxicity in TG 420 (Sewell et al., 2024) and support its use for classification and labelling.

In May 2025, the NC3Rs held online meetings with global experts, including representatives from industry and contract research organisations, to review current practices. These discussions revealed that lethality-based guidelines are still widely used, with regional preferences influencing guideline selection. For oral studies, TG 423 is preferred for REACH and biocidal regulations, while TG 425 is used for agrochemicals. For inhalation, TG 403 remains common in the USA, South America, and Korea, while TG 436 is preferred in the EU.

TG 403 also covers Concentration x Time (CxT) studies for gases, vapours, and aerosols, and aligns with US EPA guideline OPPTS 870.1300. All of these preferred guidelines still use death as the endpoint, underscoring the need for further progress.

Challenges and Opportunities

Despite progress, several barriers remain to wider adoption of evident toxicity, including concerns about regulatory acceptance, classification cut-offs, the need for point estimates of LD50/LC50, cost, and perceived uncertainty compared to mortality endpoints.

To address these challenges, the NC3Rs is working to:

- Encourage global harmonisation of guideline use, ideally selecting a single guideline per route, using the most refined methods (i.e. evident toxicity) where possible.
- Understand the drivers for the use of TG 403 and work towards deletion.
- Provide further guidance on defining and applying evident toxicity within TG 420. A proposal for this work was submitted to the OECD in 2023. However, since some member countries requested additional evidence, this is not yet on the OECD workplan. The NC3Rs is now doing

further analysis to validate the use of evident toxicity for retrospective classification and labelling, demonstrating that TG 420 does not result in misclassification.

Intended Outcomes

The NC3Rs aims to phase out mammalian acute toxicity tests that use lethality as an endpoint. This will be achieved by:

- Promoting the use of evident toxicity across relevant sectors.
- Supporting the development and regulatory acceptance of non-animal methods.
- Influencing international guideline revisions to embed refinement and replacement approaches.

Conclusion

Through scientific leadership, collaboration, and regulatory engagement, the NC3Rs is building a robust evidence base to support the refinement of acute toxicity testing and is ultimately working towards the replacement and/or elimination of these tests. This work supports a transition to more humane, scientifically robust approaches and contributes to international efforts to modernise chemical safety assessment frameworks while maintaining high standards of human health protection. The NC3Rs has led the development and validation of evident toxicity as a refined endpoint, helping to reduce animal suffering while maintaining scientific integrity, and laying the groundwork for updating OECD test guidelines.

References

Chapman K et al. (2010) The value of acute toxicity studies to support the clinical management of overdose and poisoning: a cross-discipline consensus. Regulatory Toxicology and Pharmacology 58(3): 354-359. doi: 10.1016/j.yrtph.2010.07.003

Robinson S and Chapman K (2009) Are acute toxicity studies required to support overdose for new medicines? Regulatory Toxicology and Pharmacology 55(1): 110. doi: 10.1016/j.yrtph.2009.06.012

Robinson S et al. (2008) A European pharmaceutical company initiative challenging the regulatory requirement for acute toxicity studies in pharmaceutical drug development. Regulatory Toxicology and Pharmacology 50(3): 345-352. doi: 10.1016/j.yrtph.2007.11.009

Sewell *et al.* (2015). A global initiative to refine acute inhalation studies through the use of 'evident toxicity' as an endpoint: Towards adoption of the fixed concentration procedure. *Regulatory Toxicology & Pharmacology.* 73(3):770-9. doi: 10.1016/j.yrtph.2015.10.018.

Sewell *et al.* (2018). An evaluation of the fixed concentration procedure for assessment of acute inhalation toxicity. *Regulatory Toxicology & Pharmacology.* 94:22-32. doi: 10.1016/j.yrtph.2018.01.001.

Sewell *et al.* (2024). New supporting data to guide the use of evident toxicity in acute oral toxicity studies (OECD TG 420). *Regulatory Toxicology & Pharmacology.* 146:105517. doi: 10.1016/j.yrtph.2023.105517

B) UK NC3Rs report – Reconsidering the use of two species for repeated dose toxicity testing in the development of pharmaceuticals

Introduction

Complementary to the recommendations given in chapters 4.4.1 and 4.4.2 on how to avoid repeated dose toxicity testing in a second species for plant protection products and others, similar recommendations have been worked on (and are still ongoing) for repeated dose toxicity testing in pharmaceuticals by an international multi-stakeholder collaboration under the leadership of the UK's National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs).

Insights from this project are relevant for all actions under the roadmap to avoid testing in a second species.

This chapter outlines the work of the NC3Rs to reassess the regulatory requirement for two species in regulatory toxicology studies for the development of new pharmaceuticals. This reassessment is not only critical for reducing animal use, in line with the European Commission's ambition to phase out animal testing where scientifically justified, but also for improving the efficiency and relevance of safety testing by harnessing advances in science and technology.

Background

Regulatory repeated dose toxicology studies for pharmaceuticals have traditionally required testing in both a rodent (e.g. rat or mouse) and a non-rodent (e.g. dog, minipig or non-human primate) species to support first-in-human (FIH) trials and subsequent clinical development. This two-species requirement was established over 30 years ago in International Council for Harmonisation (ICH) guidelines such as ICH M3(R2), with the aim of ensuring robust hazard identification and protecting human safety. However, scientific advances, the growing availability of safety data, and evolving regulatory perspectives now support a critical reassessment of this default requirement. New drug modalities, human-specific targets and the development of advanced *in vitro* and *in silico* technologies offer opportunities to reduce animal use without compromising human safety. In some cases, regulatory requirements have already evolved to reflect scientific progress, for example, ICH S6(R1) – the specific guideline for biotechnology-derived pharmaceuticals (i.e. biologics) – allows the use of a single species for later development studies *if* "similar" toxicities are identified in two species in First In Human-enabling studies[2].

The NC3Rs, in collaboration with the ABPI (Association of the British Pharmaceutical Industry) and global stakeholders, has led two major data-sharing initiatives to evaluate the scientific justification for the two-species requirement and explore opportunities where a single species approach might be appropriate, in alignment with the principles of the 3Rs (Replacement, Reduction, Refinement).

NC3Rs Two Species project: Phase I (2016-2020)

The first phase of the NC3Rs Two Species project brought together 37 organisations, including pharmaceutical and biotechnology companies, contract research organisations, academic institutions, and regulatory bodies. The group collected and reviewed data from 172 compounds, covering both biologics and small molecules at various stages of development (Prior et al., 2020).

The analysis revealed that, for biologics, similar toxicities were often observed in both rodent and non-rodent species during short-term studies. This finding supports the use of a single species in later development, as permitted under ICH S6(R1). However, the option to reduce to one species was not consistently applied in practice, which may reflect uncertainty around when this approach is scientifically justified and whether it will be accepted by regulators. To address this, follow-on work focused on clarifying the definition of "similar" toxicities to encourage broader uptake of the single-species approach (Prior et al., 2022). Supporting initiatives, such as the NC3Rs Virtual Second Species CRACK IT Challenge (see section 6), are also being developed to increase confidence in this approach by providing additional insight from predictive modelling of a second species alongside *in vivo* data from one species. This could help bridge the gap between traditional two-species studies and more flexible, science-driven approaches to regulatory decision-making.

The data also provided preliminary evidence that a single-species approach could be feasible for other types of drugs (e.g. small molecules), which are currently subject to the two-species requirement under ICH M3(R2). However, the evidence base was not yet sufficient to support regulatory change, highlighting the need for further investigation.

NC3Rs Two Species project: Phase II (2023-present)

Building on the findings of Phase I, a second phase was launched in 2023 with a new international working group comprising 39 organisations. This phase focuses on generating a more robust evidence base to inform regulatory decision-making for small molecules developed under ICH M3(R2), aiming to evaluate the feasibility and scientific justification for using a single species in long-term toxicity studies. Key questions include:

- How often are similar toxicities identified in short-term studies across species?
- How frequently do new or unexpected toxicities arise in long-term studies, and how do these differ between rodents and non-rodents?
- What is the theoretical risk to humans if only one species is used for long-term studies?

Data were collected from 17 companies consisting of 75 molecules across therapeutic areas such as neurology, anti-infectives, and oncology. For each molecule, information was gathered on at least two FIH-enabling and two later development studies, including study design, observed toxicities (severity, monitorability, reversibility), and their impact on decision-making.

The analysis showed that toxicity findings occurred in approximately 90% of studies, but only around 30% were considered relevant to human safety, and fewer than 20% had an impact on the clinical programme. Retrospective review indicated that, for 75% of the molecules, the same development decisions could have been made using only one species.

To explore the feasibility of prospective decision-making, a series of blinded exercises were conducted. Experts reviewed early and FIH-enabling data without access to later findings, simulating real-world decision-making. These exercises suggested that, for approximately 70% of the molecules, a single-species approach would have been appropriate, with no human-relevant toxicities missed in the absence of a second species.

Further analysis is ongoing to identify common factors in molecules where the decisions were straightforward and to explore what additional data might be needed in more complex cases where the decision was more challenging or where expert opinions differed. This work is intended to inform future regulatory guidance and support a more flexible, evidence-based approach to species selection.

NC3Rs Two Species project: Next steps

Preliminary findings from Phase II suggest that a single-species approach may be sufficient in many contexts for small molecules, particularly when supported by mechanistic understanding and Non-Animal Methods. The next phase of work will focus on two streams:

1. For molecules for which the decision to use a single species/two species was straightforward, the next phase will identify common contributing factors and explore how these could be evaluated in practice by developing tools – such as decision trees or weight of evidence frameworks – to support species selection.
2. For more complex cases, the project will explore what additional information, such as prior experience with similar molecules or the human relevance of observed toxicities, could support decision-making.

A workshop in Autumn 2025 focused on how these findings could inform future recommendations, and what evidence would be required to support regulatory change, to work towards a more flexible, scientifically justified approach to species selection, that reduces animal use without compromising human safety. The final project report will be published in 2026, with updates available via the NC3Rs Two Species project page.

NC3Rs CRACK IT Challenge: Virtual Second Species (2022-present)

To complement the data-driven approach, the NC3Rs is also supporting other activities to promote alternative approaches that could increase confidence in the use of a single *in vivo* species as part of pharmaceutical development. For example, through its open-innovation scheme CRACK IT, the Virtual Second Species project aims to develop a machine-learning-aided, multi-scale modelling framework to predict toxicological endpoints in dogs. The goal is to ultimately replace the use of dogs in chronic toxicity studies, but such a model could be used in early discovery to screen for potential 'show-stopper' toxicities in the dog, helping to focus investigations on target organs of concern (using other *in vitro* and *in silico* models) before any animal testing is conducted. In future, this approach could also support the development of virtual models for other species.

The model is designed to provide mechanistic insight and predictive toxicology data, supporting regulatory decision-making and reducing reliance on animal studies. This initiative complements the Two Species project by providing predictive tools that could further support single-species approaches in regulatory decision-making.

Intended Outcomes

The NC3Rs Two Species project aims to reduce the use of animals in regulatory toxicology studies for pharmaceuticals without compromising human safety. By identifying contexts where a single species may be sufficient, the project seeks to:

- Inform updates to international regulatory guidelines, including ICH M3, and support opportunities already allowed within ICH S6.
- Support the development of decision-making tools to guide species selection.
- Promote broader acceptance of scientifically justified alternatives.

Conclusion

Through international collaboration, data sharing, and scientific innovation, the NC3Rs Two Species project has generated a robust evidence base to support a more flexible, risk-informed approach to species selection in regulatory toxicology. The findings demonstrate that, in many cases, a single-species approach can provide sufficient information to ensure human safety, particularly when supported by mechanistic understanding and Non-Animal Methods.

This work lays the groundwork for updating international regulatory guidelines, including ICH M3, and for broader implementation of existing flexibilities under ICH S6. It also highlights the importance of developing practical tools – such as decision trees and weight-of-evidence frameworks – to support prospective decision-making.

By enabling scientifically justified reductions in animal use, the project contributes to the European Commission's ambition to phase out animal testing and advances global efforts to modernise regulatory science while maintaining high standards of human health protection. As the evidence base continues to grow, the NC3Rs remains committed to working with international partners to translate these findings into meaningful regulatory change.

References

[1] Prior *et al.*, 2020. "Opportunities for use of one species for longer-term toxicology testing during drug development: A cross-industry evaluation," *Regul. Toxicol. Pharmacol.*, vol. 113, p. 104624, Jun. 2020, doi: 10.1016/j.yrtph.2020.104624.

[2] Prior *et al.*, 2022. "Exploring the Definition of 'Similar Toxicities': Case Studies Illustrating Industry and Regulatory Interpretation of ICH S6(R1) for Long-Term Toxicity Studies in One or Two Species," *Int. J. Toxicol.*, vol. 41, no. 3, pp. 171–181, May 2022, doi: 10.1177/10915818221081439.