



An analysis of the limitations and uncertainties of in vivo developmental neurotoxicity testing and assessment to identify the potential for alternative approaches

Paparella, Martin; Bennekou, Susanne Hougaard; Bal-Price, Anna

Published in:
Reproductive Toxicology

Link to article, DOI:
[10.1016/j.reprotox.2020.08.002](https://doi.org/10.1016/j.reprotox.2020.08.002)

Publication date:
2020

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Paparella, M., Bennekou, S. H., & Bal-Price, A. (2020). An analysis of the limitations and uncertainties of in vivo developmental neurotoxicity testing and assessment to identify the potential for alternative approaches. *Reproductive Toxicology*, 96, 327-336. <https://doi.org/10.1016/j.reprotox.2020.08.002>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Journal Pre-proof

An analysis of the limitations and uncertainties of in vivo developmental neurotoxicity testing and assessment to identify the potential for alternative approaches

Martin Paparella, Susanne Hougaard Bennekou, Anna Bal-Price



PII: S0890-6238(20)30190-8

DOI: <https://doi.org/10.1016/j.reprotox.2020.08.002>

Reference: RTX 8017

To appear in: *Reproductive Toxicology*

Received Date: 3 April 2020

Revised Date: 27 June 2020

Accepted Date: 5 August 2020

Please cite this article as: Paparella M, Bennekou SH, Bal-Price A, An analysis of the limitations and uncertainties of in vivo developmental neurotoxicity testing and assessment to identify the potential for alternative approaches, *Reproductive Toxicology* (2020), doi: <https://doi.org/10.1016/j.reprotox.2020.08.002>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

An analysis of the limitations and uncertainties of *in vivo* developmental neurotoxicity testing and assessment to identify the potential for alternative approaches

Martin Paparella^{a*}, Susanne Hougaard Bennekou^b, Anna Bal-Price^c

^aDivision of Medical Biochemistry, Biocenter, Medical University of Innsbruck, Innsbruck, Austria;

^bNational Food Institute, Technical University of Denmark, Copenhagen, Denmark

^cEuropean Commission, Joint Research Centre (JRC), Ispra, Italy

*Corresponding author: martin.paparella@i-med.ac.at

Highlights:

- Uncertainty of regulatory *in vivo* DNT testing and assessment is systematically analyzed.
- A hypothesis is provided for potential improvements by alternative approaches.
- Similarities in uncertainty from *in vivo* tests and alternative approaches are discussed.
- The discussion relates to the use of alternative methods within fit-for-purpose IATAs.
- This analysis can facilitate decision making for regulatory acceptance of alternatives.

Abstract:

Limitations of regulatory *in vivo* developmental neurotoxicity (DNT) testing and assessment are well known, such as the 3Rs conflict, low throughput, high costs, high specific expertise needed and the lack of deeper mechanistic information. Moreover, the standard *in vivo* DNT data variability and in the experimental animal to human real life extrapolation is uncertain. Here, knowledge about these limitations and uncertainties is systematically summarized using a tabular OECD format. We also outline a hypothesis how alternative, fit-for-purpose Integrated Approaches to Testing and Assessment (IATAs) for DNT could improve current standard animal testing: Relative gains in 3Rs compliance, reduced costs, higher throughput, improved basic study design, higher standardization of testing and assessment and validation without 3Rs conflict, increasing the availability and reliability of DNT data. This could allow a more reliable comparative toxicity assessment over a larger proportion of chemicals within our global environment. The use of early, mechanistic, sensitive indicators for potential DNT could better support human safety assessment and mixture extrapolation. Using kinetic modelling ideally these could provide - eventually context dependent - at least the same level of human health protection. Such new approaches could also lead to a new mechanistic understanding for chemical safety, permitting determination of a dose that is likely not to trigger defined toxicity traits or pathways, rather than a dose not causing the current apical

organism endpoints. The manuscript shall motivate and guide the development of new alternative methods for IATAs with diverse applications and support decision-making for their regulatory acceptance.

Abbreviations: 3Rs, Reduce, Refine, Replace; ADME, Absorption, Distribution, Metabolism, Excretion; AO, Adverse Outcome; AOP, Adverse Outcome Pathway; BMD, Benchmark Dose; BMR, Benchmark Response; DNT, Developmental Neurotoxicity; ED₁₀, Effective Dose 10%; EFSA, European Food Safety Agency; GHS, Globally Harmonized System; hiPSCs, human induced Pluripotent Stem Cells; IATA, Integrated Approach to Testing and Assessment; KE, Key Event; LD₅₀, Lethal Dose 50%; LOAEL, Lowest Observed Adverse Effect Level; MIE, Molecular Initiating Event; MoA, Mode of Action; NOAEL, No Observed Adverse Effect Level; OECD, Organisation for Economic Co-operation and Development; PBK model, Physiology Based Kinetic model; PoD, Point of Departure; POPs, Persistent Organic Pollutants; (Q)SAR, (Quantitative) Structure Activity Relationships; QIVIVE, Quantitative In Vivo to In Vitro Extrapolation; SOP, standard operation procedure; T&A, Testing and Assessment; TG, Test Guideline;

Keywords: Uncertainties, Chemical Safety, Developmental Neurotoxicity, animal tests, validation, alternative methods, fit-for-purpose IATA

1 Introduction

1.1 Alternative approaches for DNT testing - need, availability and concepts for use

Recently, developmental neurotoxicity (DNT) scientists from academia, regulatory bodies and industry, across the world, reached an overall consensus, that current data requirements for *in vivo* DNT testing are not sufficient to adequately screen and characterize compounds potentially hazardous for the human developing brain [1, 2]. Therefore, there is a pressing need for developing alternative methods that can more rapidly and cost-effectively support the identification and characterization of chemicals with DNT potential. Alternative DNT testing methods have been proposed, including human *in vitro* models which together with non-mammalian 3Rs models and *in silico* approaches will in the future allow the development of predictive testing strategies for human DNT effects [3], [Toxicology and Applied Pharmacology DNT special issue¹].

Given the complexity of the biological processes underlying developmental neurotoxicity, it is unlikely though that any of these alternative methods will provide a standalone solution for DNT testing. It is therefore necessary to develop Integrated Approaches to Testing and Assessment (IATA), based on the combined use of multiple sources of information. Therefore, available *in vitro* assays should be incorporated into an IATA together with other sources of information including mechanistic knowledge built in Adverse Outcome Pathways (AOPs) [4], kinetic quantitative *in vitro* to *in vivo* extrapolation models (QIVIVE) and other *in silico* methods such as read across and QSARs, but also non-mammalian 3Rs models (e.g. zebrafish embryos) as well as existing animal and human data. Such IATA development should be driven by a concise problem formulation, enabling fit-for-purpose data to answer the different regulatory needs, i.a. hazard identification/characterization or risk assessment [5].

Recently, this approach has been taken on board by the Organization for Economic Cooperation and Development (OECD), aiming to develop an OECD Guidance Document on *In vitro methods to DNT testing* in collaboration with multiple stakeholders including academic scientists, industry representatives and regulators from different countries and jurisdictions [6]. Globally, several authorities/agencies and projects have committed significant investments in data generation to support this guidance development. Roughly, a 120 compounds (pesticides, drugs, environmental chemicals etc.) are being run through an *in vitro* battery in projects supported by the European Food and Safety Authority, the Danish Environmental Protection Agency (EPA), the US-EPA, NTP/NIEHS/NIH US and The H2020-funded project "EU-ToxRisk" and on this basis IATA case studies will be developed for the purpose of the guidance.

It is proposed that a battery of *in vitro* test methods should be preferably based on human induced pluripotent stem cells-derived mixed neuronal/glial cultures, since they permit evaluation of a chemical impact on critical neurodevelopmental processes, mimicking different stages of human brain development [1]. Moreover, readiness of the existing *in vitro* DNT assays for different regulatory purposes has been recently evaluated following thirteen (semi)-quantitative criteria established by DNT experts [5]. Evaluation of 16 *in vitro* assays and zebrafish behavioural tests at early development (0-5 days post fertilisation) with respect to various regulatory uses was performed and the scoring results suggest that several assays are currently at high readiness levels.

1.2 Transparency in scientific uncertainties is key for responsible decision making

In all fields of applied science, transparency in uncertainties of data and knowledge is key for a responsible decision-making [7]. This is especially important, since assessment and decision making are usually carried out by different institutions. It guarantees the scientific integrity of the assessment on one hand and well-informed, independent decision-making on the other hand. The latter usually needs to integrate the scientific information on risk metrics with various additional, socially relevant aspects.

¹ <https://www.sciencedirect.com/journal/toxicology-and-applied-pharmacology/vol/354/suppl/C>

This principle has been increasingly acknowledged in recent years for the risk assessment and risk management of chemicals. Guidance and tools have been developed for transparent characterization of the uncertainty of chemical risk metrics, such as ratios between human exposure and human limit values [8, 9].

Similarly, the need for a transparent analysis of uncertainties of the performance metrics for alternative methods, i.e. reliability and relevance, within their validation process has recently gained recognition. Specifically, the uncertainty characterization of standard reference methods for the validation of alternative methods facilitated the acceptance of alternative methods in the field of human regulatory toxicology (see following paragraphs). The types of analysis done and their consequences may be inspiring also for the DNT field.

For instance, information on the reproducibility of Test Guidelines (TGs) in the field of eye-irritation/damage and skin sensitization sets some limits for expectable correlation between data from alternative methods and from animal reference methods [10-12].

Others analyzed how the experimental variability of acute rodent lethal dose (LD_{50}) data translates to variability of Globally Harmonized System (GHS) classification [13] and it was highlighted that from a scientific perspective a borderline-range between GHS potency categories should be established. Test results falling into this borderline range, shall be considered as uncertain due to limited reliability of any test result [14]. DNT data may lead to classification for reproductive toxicity and 10% effective dose (ED_{10}) values may inform the use of specific concentration limits for mixture classification [15].

Uncertainties of the rodent carcinogenicity study-based assessment were summarized, including quantitative and qualitative information. It was concluded *inter alia* that a fully quantitative validation is not feasible due to the high complexity and potential variability in data generation and assessment and this represents an uncertainty of the reference method as such [16]. Nevertheless, the applied assessment scheme allows a semi-quantitative and qualitative comparison of the limitations and uncertainties of current standards versus new alternative approaches. This may support a best-possible informed decision on the acceptability of the new approaches.

Information on the limitations and uncertainty of standard *in vivo* DNT testing has also been published, e.g. [5, 17, 18]. However, a comprehensive summary of all potential limitations and uncertainties, similar to the work for carcinogenicity [16] is still missing. Consequently, here such a similar summary for limitations and uncertainties is provided for the current regulatory standard *in vivo* DNT testing and assessment, following the same OECD template [19, 20].

2 Building the hypothesis: Limitations and uncertainties of *in vivo* DNT testing and assessment – can be reduced with alternative approaches?

In vivo DNT testing and assessment has a number of limitations and uncertainties and these are summarized here. Furthermore, we indicate which of these limitations and uncertainties may be reduced with alternative approaches including *in vitro* testing and *in silico* modelling

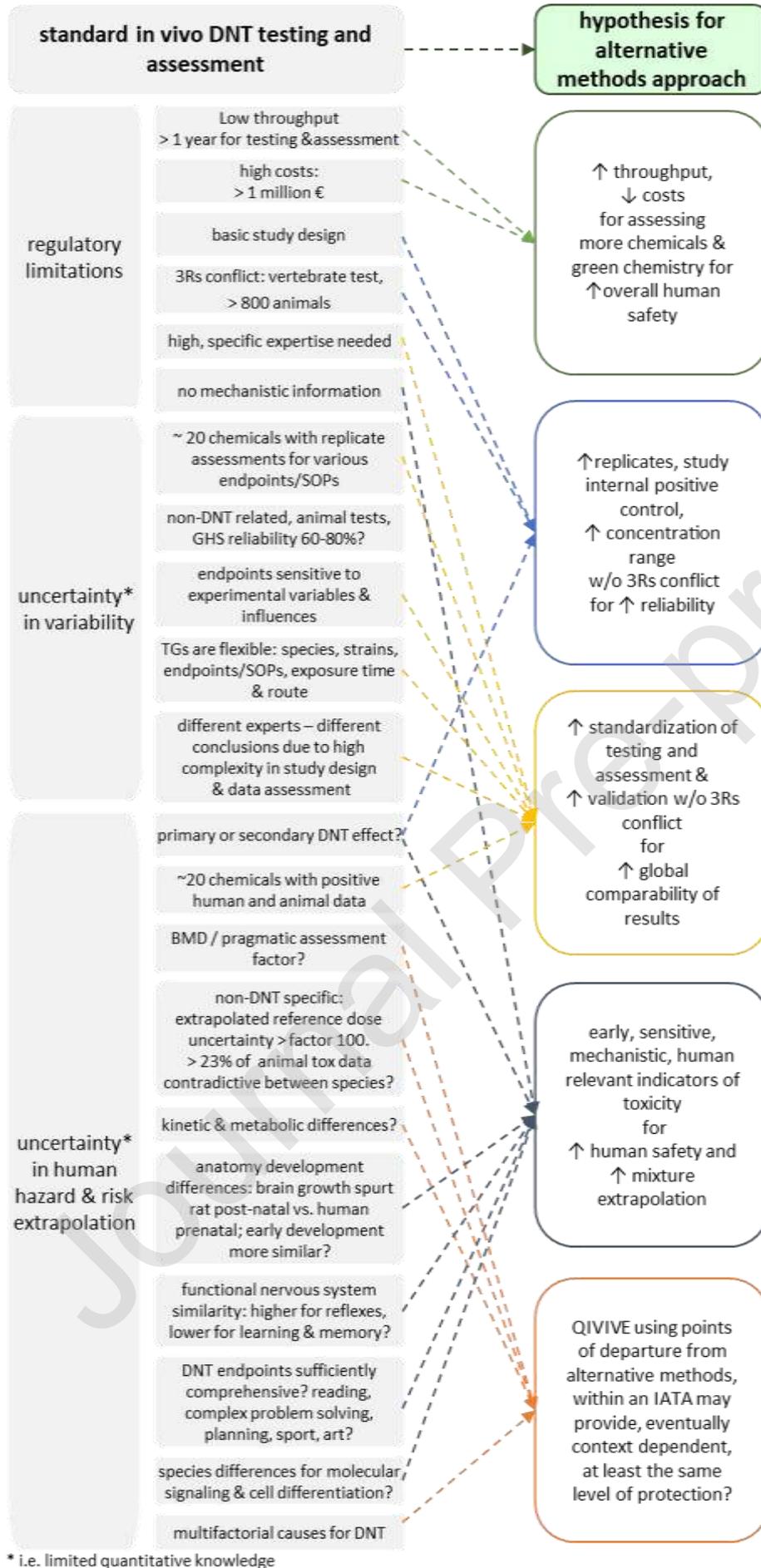
For a top level overview, the main aspects of this discussion are illustrated in Figure 1 and summarized in section 2. The figure and summary intend to provide a structure that could support a targeted discussion by regulatory experts, for the one or the other aspect that may be relevant for the development of DNT IATA(s) for diverse regulatory purposes. On the one hand, the results of such targeted discussion may motivate and guide the development of new alternative approaches to improve the current methodological toolbox. On the other hand, it shall support regulatory decision-makers, for agreeing what performance of IATAs including alternative approaches may represent an overall improvement for the various regulatory needs.

In the supplement Table S1, information about the limitations and uncertainties of the use of *in vivo* standard DNT TGs is presented within an OECD standard tabular format. This format was originally developed to characterize alternative methods as individual information sources to be used within IATAs [20]. The tabular summary was applied to carefully consider all potential limitations and

uncertainties of the use of the in vivo DNT TGs and to develop the summarizing figure 1 and text for section 2.1 to 2.3. in this manuscript. Table S1 may be further amended and refined, as far as useful, within the ongoing regulatory experts' project on DNT IATA(s) development. Applying the same systematic characterization scheme for both, the use of current DNT in vivo standards and the use of alternative methods, may facilitate decision making on the acceptability of fit-for-purpose IATA(s) for various regulatory applications. An internationally reviewed/revised version of the S1 table could also be included as an annex to the OECD IATA guidance, which is work in progress. It could serve as a reference for some specific discussions and sections within this OECD guidance.

Journal Pre-proof

Figure 1: Building the hypothesis: Regulatory limitations and uncertainties of in vivo DNT testing and assessment could be reduced with alternative approaches.



2.1 Regulatory limitations of standard in vivo DNT Testing & Assessment

The standard in vivo DNT testing and assessment currently represents as a screening method for the evaluation of many chemicals with a broad set of potentially relevant endpoints. However, this regulatory purpose is limited by several aspects:

The regulatory limitations of standard rodent in vivo DNT testing and assessment relate to its low throughput, requiring more than one year for one substance, the high costs, i.e. above 1 million Euro [21] and the conflict with the global 3Rs goals due to the use of vertebrates (Table S1, lines 5.7.).

Moreover, a high level of specific expertise in conduct, assessment and DNT data interpretation is needed, requiring continuous, species- and method- specific (re)training. This limits the availability of experienced experts, especially since –due to the many other limitations- the frequency of in vivo DNT testing and assessment is low [5, 22, 23] creating a negative feedback circuit (Table S1, lines 5.7.).

Furthermore, due to the 3Rs conflict [24], practicalities and costs, the basic study design is limited in terms of feasible animal numbers. This restricts the minimal effect size that can be detected as statistically significant. Minimal detectable effects size estimates are available, but these vary considerably (and include values beyond 100% effect size; [25-28], Table S1 line 5.4.2.3). Therefore, as true for all in vivo assessments, standard in vivo DNT data assessment requires consideration of biological relevance in addition to statistical significance and this increases uncertainty in the data assessment due to complexity, e.g. [29, 30]. Moreover, negative historical controls need to be considered in this context, but this is hampered due to low frequency of testing, the heterogeneity of test designs and large flexibility in data interpretation. The limited animal number may also lead to testing of high doses, findings of indirect DNT effects and missing non-monotonic-dose-response relationships [25, 31, 32]. Further uncertainty relates to the lack of study internal positive controls for study validity assessment [33] (Table S1, lines 5.7.).

Last, but not least current standard in vivo DNT testing does not provide mechanistic information, supportive for the identification of early sensitive indicators of potential DNT effects and suitable for mixture toxicity investigations [34, 35] (Table S1, lines 5.7.).

2.2 Uncertainties in variability of in vivo DNT Testing and Assessment

There is uncertainty related to knowledge about the variability of standard in vivo DNT data. Replicate testing and assessment is available for about 20 chemicals [36]. Estimates for variability vary, depending on the chemical, the endpoints and the approach to generate the variability estimate (available estimates include similarity values below a similarity proportion of 0.5 from 1 [37, 38]. Of course, the estimate cannot be comprehensive for all DNT endpoints and related methods. Published, non-DNT related estimates for reliability of animal test data appear to be in the range between 60%-70% for a three category GHS classification and 70-80% for a two category GHS classification (for skin sensitization and eye irritation/damage [10-12, 39-41] or about 57% for carcinogenicity [42]. No internationally agreed quantitative variability estimate is available for DNT [36]. By default may we expect lower variability for the very complex DNT study endpoints (Table S1, lines 5.8.)?

However, experimental variability is expected. Behavioral endpoints are reported to be sensitive to several experimental variables (e.g. for motor activity, the shape, size, movement detection system and related activity metrics and data processing). Endpoints are also sensitive to potential external influences to the experiment (e.g. noise, light, odor, handling, time of observation relative to light-dark cycle and dosing, age, test history, stress), to individual animal variability and to some subjective measurements (e.g. clinical observations, histopathology). Some guidance is available to limit some of this variability (Table S1 lines 5.3.) [25].

There are some differences between the in vivo DNT test guidelines. For example, in contrast to the OECD TG 426 [43], the OECD TG 443 [44] includes pre-mating exposure, but uses fewer pups for testing and includes cognitive learning and memory tests only if triggered based on available data. Therefore, OECD TG 443 may be more comprehensive in terms of the exposure regime, but less comprehensive in terms of potentially relevant DNT endpoints covered and overall conclusions may

differ between tests from different guidelines. Moreover standard test guidelines allow a great deal of flexibility in terms of endpoints and related methods, exposure time, exposure route and top dose selection. This allows the testing laboratory or registrant to use the endpoints most appropriate for the specific chemical, the related toxicological knowledge and regulatory need. However, in regulatory practice, information to select the best study design may often be very limited and different study directors may decide for different study designs. An informed hypothesis-based targeted testing is rarely possible. Therefore, the guideline inherent flexibility limits comparisons of toxicity between studies and chemicals (Table S1, lines 2.1.).

It is a highly complex task to select a specific study design in terms of animal strain, endpoints and related methods, exposure periods and exposure routes as well as top-doses (Table S1 lines 5.3.). In addition, data analysis and interpretation requires the need to consider various main effects, interactions, data-dependencies, uncertainties, statistical methods and integration of statistical significance and biological relevance. Often data from more than one study need to be assessed, which may be conflicting (Table S1 lines 5.4.). Taken together this is a complex task and different expert groups may come to various scientifically legitimate conclusions. Data informing on a realistic probability for such situations of ambiguity are available for non-DNT related regulatory toxicology assessments [29, 30, 45] and include a chance of 40% and more for different expert assessments (Table S1 line 4.12.11).

2.3 Uncertainties in human hazard and risk extrapolation

Some uncertainty relates also to the extrapolation from the standard rodent in vivo DNT testing to human hazard and risk estimates. For about 20 chemicals, positive human and at least high dose positive standard animal DNT data are available, indicating the potential to predict human DNT with standard animal approaches [23, 46, 47]. Of course this does neither mean that all chemicals with human DNT hazard can be identified with standard animal DNT testing and assessment, nor that all identified animal DNT hazard is relevant for humans (see Table S1 lines 5.9.).

Quantitative extrapolation from the experimental Benchmark-Dose(BMD) or No-Observed-Adverse-Effect-Level (NOAEL) to a human reference value may contain high uncertainty especially in the absence of specific kinetic and metabolism (ADME) data (according to [46], up to a factor of 10.000, see table S1, line 5.9.1.2). It is also noted that estimating the exposure and ADME in pups and extrapolating this to the human situation may contain very significant uncertainties [33, 48, 49] (Table S1, lines 4.3.2, 4.12.4-7, 5.1.1.4, 5.1.1.8).

Applying the usual pragmatic standard assessment factors (often 100) to BMDs or NOAELs results in human reference values with an unknown protection level (unknown in terms of probability for population fractions under risk for the critical effect and the related uncertainty). However, using data-based probabilistic extrapolation models human reference values are estimated with a 5th/95th percentile range that spans at least a factor of 100 for a protection target of $\leq 1\%$ population under risk. This uncertainty range becomes unknown for higher protection targets of e.g. $\leq 0.0001\%$ population under risk due to limitations in the data available for modelling. It is also noted that the current database for this probabilistic model does not include DNT specific data [9]. Furthermore, there are qualitative differences between species that are not captured by the probabilistic assessment factors (as summarized in [50] about 23% of data sets; see Table S1 line 5.9.1.3.).

Extrapolation uncertainties may also relate to the fact that primary DNT effects sometimes may not be distinguished from secondary, non-specific consequences of maternal and other toxic effects (e.g. acute neurotoxicity or strongly reduced feed/water intake or general systemic toxicity leading to strong maternal body weight decrease, reduced litter size & weight and slower postnatal development). In this case, the DNT effects are pragmatically considered to be of concern [25], though the relevance for expectable human hazard in the real world situations remain uncertain (see Table S1 lines 2.2.3).

In qualitative terms, extrapolation uncertainties relate to rat to human differences in brain development (morphology and functional anatomy). While the general processes and sequence in brain development are conserved between species, the major brain growth spurt is postnatal in rat, but prenatal in human. Thus exposure route, related kinetics and metabolism differ for that critical

period, i.e. milk and/or feed for rats versus placenta for human [33, 49]. It is reported that the similarity of the functional nervous system is higher for reflexes but lower for learning and memory [25] – and other higher cerebral tasks (like reasoning, reading, planning, organizing) and complex nervous system functions (like advanced sportive or artistic activity) cannot be tested in animals as such. Given the complexity of neurodevelopmental disorders, e.g. *autism* which prevalence dramatically increased during the recent years [51, 52], the use of animals causes significant uncertainties since they do not develop the full spectrum of features characteristic for such diseases [53, 54]. Some differences were also reported at the level of molecular signaling and cell differentiation [55-58]. Currently, it remains an unanswered question how comprehensive the current standard animal DNT testing and assessment is for relevant human neuronal functionalities (Table S1 line 5.1.1.).

Last, but not least, DNT may be the result of many causes, i.e. a combination of epigenetic and genetic background, socio-economic status, diet, life style, stress and co-exposure including environmental contaminants and drugs or maternal infection and viruses, for instance [59]. The dose at which chemicals may contribute to DNT will depend on these other real-world modulators, which currently cannot be covered by any testing (Table S1 line 4.12.1).

2.4 Building the hypothesis: How could alternative approaches reduce current regulatory limitations and uncertainties?

Recently alternative models for DNT testing and assessment have become available. For instance, mixed neuronal/glial cultures derived from human induced pluripotent stem cells (hiPSCs) can be used as they allow evaluation of chemical impacts on key neurodevelopmental processes, by reproducing different windows of exposure during human brain development [1]. If these processes (e.g. cell proliferation, migration, synaptogenesis, neuronal network formation and function) are impaired as a result of chemical exposure, they can be assessed in a quantitative manner using *in vitro* assays and serve as reliable readouts for DNT effects. Recently, the European Food Safety Agency (EFSA) has published a detailed report on the evaluation of the currently available DNT *in vitro* test methods, including human models, concluding that a variety of *in vitro* methods, covering early and late stages of neurodevelopment are already available and could be used to predict DNT effects [60]; for a summary of methods and concepts, see also section 1.1.).

2.4.1 Potentially reduced limitations and uncertainties?

Such alternative approaches could reduce several of the current regulatory limitations, due to their relevance to human biology, better mechanistic understanding of toxicity and potential for much higher throughput and strongly reduced costs and coherence with the global 3Rs goals. This could permit to assess many more chemicals, eventually nanomaterials, mixtures and environmental media and allow assessing potentially less problematic, including green chemistry, early in chemical development [61]. Toxicological data may also be updated more frequently by retesting of chemicals in line with the scientific progress. With such an expansion of toxicological data, the overall human safety could be increased.

Moreover, with alternative approaches, a higher number of replicates, broader concentration ranges and study internal positive controls could be tested without 3Rs conflict. In addition, alternative testing and assessment approaches are usually standardized to a higher degree. Altogether, this could reduce the variability of data and the “biodiversity” of expert assessments. A parallel assessment of statistical and biological relevance a posteriori to testing may not be necessary for alternative approaches. Due to the absence of a 3Rs conflict with retesting, the improved standardization may also facilitate better validation for reproducibility of specific study designs. A higher data reliability could also increase the sensitivity of the method and thereby favor the distinction between primary and secondary DNT effects. Because of all these advances, the comparability of data generated for similar or different chemicals could be increased and this would translate to more reliable toxicity comparisons between chemicals, more reliable classifications and

more reliable human reference values. This would be of advantage for global trade and global regulation of the more than 100.000 chemicals on the market².

2.4.2 Potentially new uncertainties?

Alternative approaches may test for early, sensitive, mechanistic indicators of toxicity. It may be that this introduces some new uncertainties. However, these potentially new uncertainties are **conceptually** not different from current in vivo based testing and assessment (Table 1).

Table 1: Potential new uncertainties from alternatives are conceptually similar to uncertainties from current standard in vivo testing and assessment

Alternative testing and assessment	Current standard in vivo testing and assessment
Early events are “just” indicators for potential toxicity. They may be compensated or not. The knowledge about the complexity of regulation at organism and population level is limited.	Standard rodent in vivo DNT studies may engage groups of endpoints, i.e. for clinical signs, motor activity, motor & sensory function, learning and memory, histopathology. For study interpretation not just long lasting, obvious neurological deficits, but all types of effects are considered as indications for DNT, including transient, time-frame specific effects, increasing or decreasing, and also changes in variability ([25]; Table S1, lines 4.9.1., 5.4.1.1., 5.4.1.4). Moreover, the higher cerebral tasks (like reasoning, reading, planning, organizing) and complex nervous system functions (like advanced sportive or artistic activity) are not tested in animals. TG443 does not even test for learning and memory (Table S1 line 2.2.4). Socio-economic status, diet, life-style and other environmental influences (which currently cannot be tested with any system) may compensate or increase neurological deficits in real human life [59] (Figure 2, Table S1 line 4.12.1). Thus, the tested standard in vivo endpoints are considered “just” as coarse indicators for a more general neurological system health status [62].
AOPs rely on a broad mechanistic knowledge coming from many diverse data sources. Both (AOPs and basic science) are in continuous development.	Aligning the rodent DNT endpoints (for motor activity, startle response, conditioned fear, various water mazes) with functionally similar neuropsychological effects that can be tested and observed in humans is ongoing work (Table S1, line 5.1.1.9).
The early mechanistic events tested may not cover all potential mechanisms leading to DNT.	The issue of comparability of human and rodent cognitive measures in studies of pharmacology and neurotoxicology has generated a large body of literature [53, 63-65]. How comprehensive is the mechanistic coverage of the current standard in vivo approach? (Table S1, lines 5.1.1.5., 5.1.1.9)
Therefore, “positive” data may be more reliable than “negative” data.	Given the many potentially relevant variants of standard in vivo DNT test designs (in terms of endpoint selection, testing and assessment approaches), “positive” in vivo DNT data are more reliable than “negative” in vivo DNT data.
In vitro metabolism is limited to the biotransformation capacity of the isolated test system and may be dissimilar to in vivo metabolism.	Metabolism may vary between species and environmental influences. How much do we know about this variability? (Table S1, lines 5.5., 5.1.1.8; [48])
Defining a critical benchmark-response (BMR) within alternative methods, to be used as Point of Departure (PoD)	Currently, in regulatory practice the discussion of critical effect sizes (as BMR for PoD derivation) is secondary to consideration of statistical significance, i.e. NOAELs/LOAELs are used in a pragmatic way. Moreover, critical effect sizes are not harmonized at

² See e.g. ECHA Classification and Labelling inventory, i.e. about 142,000 should be on the European market.

<p>for limit value derivation, may require to establish a new convention.³</p>	<p>regulatory level for the various in vivo endpoints. Often the BMR₁₀ is used by default. However, also the latter may be disputed, at least for quantal data: We may consider that the dose-response relationship in the animal-test is very much influenced by experimental conditions including species/strain selection and therefore does not contain really relevant information. Theoretically, we may be more interested to know the dose that causes the effect in an average animal. In fact all experimental standardization aims to reduce variability between animals. In this case, for quantal data, the BMR₅₀ would be more relevant. Human variability will in any case be modelled separately based on human data [9].</p> <p>In summary, defining the critical effect size for in vivo endpoints, needs convention.</p>
<p>Testing and/or modelling in vitro kinetics as well as QIVIVE modelling is necessary for translating in vitro results to human in vivo reference values. This introduces measuring and modelling uncertainty⁴.</p>	<p>Kinetic inter-species extrapolation uncertainties may be particularly uncertain for reproductive toxicity studies [33, 48, 49] (Table S1, lines 4.3.2, 4.12.4-7, 5.1.1.4, 5.1.1.8).</p> <p>The current application of standard assessment factors to in vivo derived BMDs or NOAELs/LOAELs results in reference values with an unknown protection level (in terms of population fraction under risk and related uncertainty).</p> <p>Probabilistic extrapolation models are needed to extrapolate from the rodent tests to humans [9], which introduces modelling uncertainties (see Table S1, line 5.9.1.3).</p>

In summary, effects at the molecular and cellular level may be the earliest testable biomarkers or key events that may or may not progress to an adverse outcome at the organism and population level, depending on many biological factors that cannot be tested with any in vivo or in vitro system (Figure 2). In this context, the AOP framework has been recognized in regulatory science as an efficient and effective tool for capturing existing knowledge describing the linkage between mechanistic information at these different biological levels and adverse outcome (AO) of regulatory relevance.

From that perspective, alternative approaches could allow the most precautionary testing approaches that could also be calibrated to a desired sensitivity. Moreover, the increased data reliability and the mechanistic information gained from alternative approaches could provide several other advantages: The distinction between primary and secondary DNT effects could be better supported. Extrapolation from individual chemical testing to mixture toxicity assessment could become better informed.

Recently the AOP concept has been applied for the assessment of developmental neurotoxicity induced by mixtures of Persistent Organic Pollutants (POPs) [35] and mixture of metals [66]. These studies showed that identified common key events within a network of AOPs relevant to DNT can serve as reliable and robust anchors, for in vitro assays selection for DNT testing, triggered not only

³ For understanding, how much the selection of a BMR₅ versus e.g. a BMR₅₀ matters we may consider: On average, how high is the difference between selecting the corresponding concentrations (i.e. BMC₅ or BMC₅₀) as PoDs? If expressed as a probability distribution over a larger set of chemicals, what is the proportion of this uncertainty distribution relative to the total uncertainty of the human reference value after QIVIVE and intra-species variability modelling?

⁴ Furthermore it is noted that QIVIVE modelling is work in progress at OECD level and it may take several years for their regulatory application. Currently Guidance is in development by the OECD Working Party for Hazard Assessment aiming to: 1) Summarize a scientific workflow for characterizing and validating physiology based kinetic (PBK) models, with emphasis on the use of in vitro and in silico data for ADME parameters, and in scenarios where in vivo kinetic data are limited or unavailable to parameterize model parameters. 2) Identify knowledge sources on in vitro and in silico methods that can be used to generate ADME parameters for PBK models. 3) Develop an assessment framework for evaluating PBK models for specific purposes, with emphasis on the major uncertainties underlying the model predictions. 4) Provide a template for documenting PBK models in a systematic manner. 5) Provide a checklist to support the evaluation of PBK model applicability according to context of use. In addition to this case studies are in development

by single chemicals but also mixtures. Furthermore, mechanistic knowledge built in DNT AOPs facilitates data interpretation and their possible application for regulatory purposes. Within the currently available DNT AOPs the most frequent AO is defined as cognitive damage or learning and memory impairment in children [67].

On the one hand, uncertainties in extrapolation from animal in vivo endpoints and related BMDs or NOAELs/LOAELs to human effects and reference doses (section 2.3.) are well recognized. On the other hand, uncertainties in extrapolation from alternative methods endpoints to human organism level effects are also obvious. Considering the potential for high precaution and calibration of endpoints from alternative methods, maybe – eventually context dependent - at least a similar level of human health protection can be reached with alternative approaches?

Recently a few AOPs have been developed where impairment of learning and memory in children has been defined as adverse outcome (AO). These AOPs facilitate mechanistic understanding of disturbed signaling pathways and/or neurodevelopmental processes such as neurite outgrowth, neuronal/glial differentiation, synaptogenesis, neuronal network formation and function etc., defined as Key Events (KEs) involved in impairment of cognitive functions. These KEs can serve as anchors for in vitro assays, permitting to characterize in a quantitative manner which chemicals (at which concentration and time of exposure) could be considered as potential triggers of cascade of events leading to cognitive damage in children [35]. Such in vitro assays are, first of all, suitable for mapping a chemical-induced disturbance of signaling pathways at the cellular level and nowadays, can be performed using human models (e.g. mixed neuronal/glial culture derived from induced pluripotent stem cells) avoiding species extrapolation [1].

Current perspectives for the development of an IATA foresee a tiered strategy. After evaluating available DNT evidence, alternative approaches may be included within tier 1 and 2 assessments. Depending on the resulting data and the regulatory context this may be followed up by targeted rodent testing [1, 5], if necessary. Of course, reliability and relevance of the alternative tier 1 and 2 testing needs to be sufficiently high, i.e. at least significantly above 50%, in order to justify their use. Considering the uncertainties summarized above in the standard in vivo approach (Figure 1, sections 2.1.-2.3. and Table S1), the extra margin of certainty achievable with additional tier 3 in vivo testing, still needs to be evaluated. However, with further progress of this specific regulatory science, the current perspectives for tiered testing and assessment strategies may evolve further.

3 (How) may regulatory toxicology evolve to make better use of alternative approaches?

In principle, discussion of two questions may be helpful for the evolution of regulatory toxicology: 1) Which new approaches can be developed to serve the current regulatory toxicology system. 2) (How) may the current regulatory toxicology framework be adapted to make better use of the new alternative approaches?

While a lot of work is in progress to provide answers to the first question (see e.g. Toxicology and Applied Pharmacology DNT special issue⁵), discussions tackling the second question is available, but still scarce [16]. Therefore, one thought-starter is further spread here (Figure 2): Can we use alternative methods data as early sensitive, mechanistic indicators for a potential DNT hazard and QIVIVE modelling to derive a “probably acceptable human effect level”, in the sense that the alternative approaches may predict likely safe exposures for specific toxicity traits (e.g. no effects on key neurodevelopmental processes including cell differentiation, migration, neurite outgrowth, synaptogenesis, network function), rather than the current organism level endpoints? AOPs provide the mechanistic understanding of relationships between the molecular initiating event which triggers a cascade of key events at different biological levels (cellular, tissue, organ) and results in adverse outcome at organism level.

Different approaches provide different types of data and this may justify evolving current adversity definition and GHS classification criteria or even creating a new in vitro mode of action (MoA) hazard

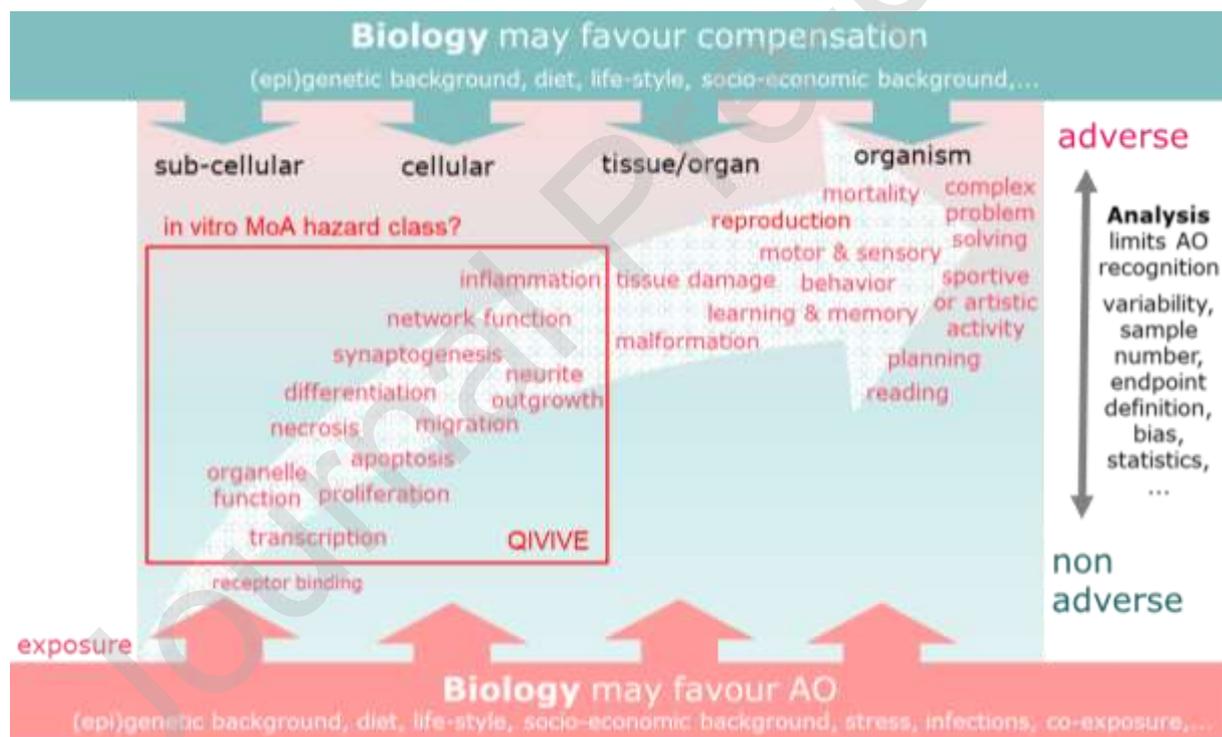
⁵ <https://www.sciencedirect.com/journal/toxicology-and-applied-pharmacology/vol/354/suppl/C>

GHS class, depending on how different the data are and if supported by substantial social benefits. QIVIVE supported subcategorization to account for potency differences could be key for future regulation.

As discussed earlier [16], it is noted that many biological modifiers (such as (epi)genetic background, diet, life-style, socio-economic background, stress, infections, co-exposure) can influence the progression of effects from the molecular/cellular level to an adverse outcome at the organism level. Therefore, the concentrations at which the effects at the molecular/cellular level can progress to organism level effects can be very variable in real life. These biological modifiers are neither assessed with in vivo methods. Therefore, quantitative modelling of the complex, potential compensatory, mode of actions may not be as important as highly reproducible test results. If molecular initiating events (MIEs) and/or KEs at the cellular/tissue level are affected this may be considered as an increased chance that AOs may develop, depending on many variable and also unknown biological modifiers; in figure 2, this is indicated by the green “non-adverse” to red “adverse” background shading. Furthermore, consider the white arrow in figure 2, which broadens from the sub-cellular to the population level, indicating that one could expect that the experimental variability is likely to increase with the complexity of the system. For alternative and in vitro (MIE, KE) based test systems this may provide the added benefit of relatively reduced uncertainty for variability compared to the more complex in vivo systems?

Several considerations of conceptual similarity of current animal testing approaches and future in vitro and in silico approaches indicate that such an evolution of regulatory toxicology may be possible (Table 1).

Figure 2: Understanding chemical safety, as a dose that is likely not to trigger defined toxicity traits or pathways, rather than the current apical organism level endpoints (adapted from [16]).



4 Conclusion

The current in vivo DNT testing and assessment approach provides a hazard characterization with relevant regulatory limitations and uncertainty. A summary of respectively available knowledge is illustrated here and provided in a systematic OECD standard-template in the supplement (Table S1). This knowledge base may be reviewed, revised, eventually updated and further developed within the

broad expertise of the OECD and EFSA DNT Working Groups. It could serve to support decision making for the acceptance of alternative methods based IATAs, since the latter should represent an overall improvement of the current approach.

Theoretically, also the in vivo DNT test guidelines as such could be improved, in terms of their basic study design, further mechanistic and quantitative validation of endpoints and a higher standardization of testing and assessment. However, such a validation would require many animals and it would be very costly in terms of time and money. Moreover, for a reliable use, historical control databases would need to be built, which would take even more time. Thus, it would be difficult to keep any standard in vivo approach up to date with scientific progress. Moreover, many of the limitations and uncertainties of the standard in vivo DNT approach, are of generic nature and also apply to other animal tests, e.g. limitations of reliability and sensitivity due to constraints for the basic test design, lack of deeper mechanistic information, uncertainties for species differences in metabolism and kinetics and limited quantitative knowledge of variability and uncertainty related to hazard and risk extrapolation (Table S1, uncertainties specified with “G” for generic or “S” for DNT specific). Last, but not least, the practical regulatory limitations in terms of testing throughput and 3Rs conflict may not be sufficiently improved considering current data needs.

Here, we outlined a hypothesis which regulatory limitations and uncertainties of the in vivo DNT test guidelines could be reduced by new alternative approaches: Gains in 3Rs compliance, reduced costs, higher throughput, improved basic study design (replicates, study internal positive controls, concentration range), permitted better mechanistic understanding of toxicity, higher standardization of testing and assessment and validation without 3Rs conflict - could increase the availability and reliability of DNT data for thousands of chemicals. This could allow a comparative toxicity assessment over a larger proportion of chemicals within our global environment, which are currently without DNT data. The use of early, sensitive, mechanistic indicators for potential DNT could support precautionous human safety assessment and mixture extrapolations. Combined with QVIVE modelling, ideally this could provide – eventually context dependent (fit-for-purpose) - at least the same level of human health protection. Such new approaches could also lead to a new understanding for chemical safety, as a dose that is likely not to trigger defined toxicity traits or pathways, rather than the current apical organism level endpoints.

This hypothesis could now be tested in upcoming work, which aims for specifying and characterizing alternative approaches and aims for analyzing how these new approaches may reduce regulatory limitations and uncertainties of current DNT assessments based on animal experiments and observational human data.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement: The work of Martin Paparella at the Medical University of Innsbruck is financed by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology, Department V/5 – Chemicals Policy and Biocides.

Disclaimer: The scientific views presented in this paper are those of the authors alone and do not necessarily reflect official views of their respective institutions.

References

- [1] A. Bal-Price, F. Pistollato, M. Sachana, S.K. Bopp, S. Munn, A. Worth, Strategies to improve the regulatory assessment of developmental neurotoxicity (DNT) using in vitro methods, *Toxicol Appl Pharmacol* 354 (2018) 7-18. 10.1016/j.taap.2018.02.008
- [2] A. Bal-Price, K.M. Crofton, M. Leist, S. Allen, M. Arand, T. Buetler, N. Delrue, R.E. FitzGerald, T. Hartung, T. Heinonen, H. Hogberg, S.H. Bennekou, W. Lichtensteiger, D. Oggier, M. Paparella, M. Axelstad, A. Piersma, E. Rached, B. Schilter, G. Schmuck, L. Stoppini, E. Tongiorgi, M. Tiramani, F. Monnet-Tschudi, M.F. Wilks, T. Ylikomi, E. Fritsche, International Stakeholder NETWORK (ISTNET): creating a developmental neurotoxicity (DNT) testing road map for regulatory purposes, *Arch Toxicol* 89(2) (2015) 269-87. 10.1007/s00204-015-1464-2
- [3] E. Fritsche, P. Grandjean, K.M. Crofton, M. Aschner, A. Goldberg, T. Heinonen, E.V.S. Hessel, H.T. Hogberg, S.H. Bennekou, P.J. Lein, M. Leist, W.R. Mundy, M. Paparella, A.H. Piersma, M. Sachana, G. Schmuck, R. Solecki, A. Terron, F. Monnet-Tschudi, M.F. Wilks, H. Witters, M.G. Zurich, A. Bal-Price, Consensus statement on the need for innovation, transition and implementation of developmental neurotoxicity (DNT) testing for regulatory purposes, *Toxicol Appl Pharmacol* (2018). 10.1016/j.taap.2018.02.004
- [4] A. Bal-Price, P.J. Lein, K.P. Keil, S. Sethi, T. Shafer, M. Barenys, E. Fritsche, M. Sachana, M.E.B. Meek, Developing and applying the adverse outcome pathway concept for understanding and predicting neurotoxicity, *Neurotoxicology* 59 (2017) 240-255. 10.1016/j.neuro.2016.05.010
- [5] A. Bal-Price, H.T. Hogberg, K.M. Crofton, M. Daneshian, R.E. FitzGerald, E. Fritsche, T. Heinonen, S. Hougaard Bennekou, S. Klima, A.H. Piersma, M. Sachana, T.J. Shafer, A. Terron, F. Monnet-Tschudi, B. Viviani, T. Waldmann, R.H.S. Westerink, M.F. Wilks, H. Witters, M.G. Zurich, M. Leist, Recommendation on test readiness criteria for new approach methods in toxicology: Exemplified for developmental neurotoxicity, *ALTEX* 35(3) (2018) 306-352. 10.14573/altex.1712081
- [6] M. Sachana, A. Bal-Price, K.M. Crofton, S.H. Bennekou, T.J. Shafer, M. Behl, A. Terron, International Regulatory and Scientific Effort for Improved Developmental Neurotoxicity Testing, *Toxicol Sci* 167(1) (2019) 45-57. 10.1093/toxsci/kfy211
- [7] SAEPA, Making sense of science for policy under conditions of uncertainty and complexity, SAEPA Evidence Review Report No. 6 (2019). 10.26356/MASOS
- [8] EFSA, Guidance on Uncertainty Analysis in Scientific Assessments, *EFSA Journal*, 2018, p. 5123.
- [9] WHO, Guidance document on evaluating and expressing uncertainty in hazard characterization, *IPCS*, 2017.
- [10] E. Adriaens, J. Barroso, C. Eskes, S. Hoffmann, P. McNamee, N. Alépée, S. Bessou-Touya, A. De Smedt, B. De Wever, U. Pfannenbecker, M. Tailhardat, V. Zuang, Retrospective analysis of the Draize test for serious eye damage/eye irritation: importance of understanding the in vivo endpoints under UN GHS/EU CLP for the development and evaluation of in vitro test methods, *Arch Toxicol* 88(3) (2014) 701-23. 10.1007/s00204-013-1156-8
- [11] J. Barroso, U. Pfannenbecker, E. Adriaens, N. Alepee, M. Cluzel, A. De Smedt, J. Hibatallah, M. Klaric, K.R. Mewes, M. Millet, M. Templier, P. McNamee, Cosmetics Europe compilation of historical serious eye damage/eye irritation in vivo data analysed by drivers of classification to support the selection of chemicals for development and evaluation of alternative methods/strategies: the Draize eye test Reference Database (DRD), *Arch Toxicol* 91(2) (2017) 521-547. 10.1007/s00204-016-1679-x
- [12] S. Hoffmann, N. Kleinstreuer, N. Alépée, D. Allen, A.M. Api, T. Ashikaga, E. Clouet, M. Cluzel, B. Desprez, N. Gellatly, C. Goebel, P.S. Kern, M. Klaric, J. Kühnl, J.F. Lalko, S. Martinozzi-Teissier, K. Mewes, M. Miyazawa, R. Parakhia, E. van Vliet, Q. Zang, D. Petersohn, Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database, *Crit Rev Toxicol* 48(5) (2018) 344-358. 10.1080/10408444.2018.1429385
- [13] S. Hoffmann, A. Kinsner-Ovaskainen, P. Prieto, I. Mangelsdorf, C. Bieler, T. Cole, Acute oral toxicity: variability, reliability, relevance and interspecies comparison of rodent LD50 data from literature surveyed for the ACuteTox project, *Regul Toxicol Pharmacol* 58(3) (2010) 395-407. 10.1016/j.yrtph.2010.08.004
- [14] M. Leontaridou, D. Urbisch, S.N. Kolle, K. Ott, D.S. Mulliner, S. Gabbert, R. Landsiedel, The borderline range of toxicological methods: Quantification and implications for evaluating precision, *ALTEX* 34(4) (2017) 525-538. 10.14573/altex.1606271
- [15] ECHA, Guidance on the Application of the CLP Criteria (Version 5.0), in: ECHA (Ed.) 2017.

- [16] M. Paparella, A. Colacci, M.N. Jacobs, Uncertainties of testing methods: What do we (want to) know about carcinogenicity?, *ALTEX* 34(2) (2017) 235-252. [10.14573/altex.1608281](https://doi.org/10.14573/altex.1608281)
- [17] L. Smirnova, H.T. Hogberg, M. Leist, T. Hartung, Developmental neurotoxicity - challenges in the 21st century and in vitro opportunities, *ALTEX* 31(2) (2014) 129-56. [10.14573/altex.1403271](https://doi.org/10.14573/altex.1403271)
- [18] A. Terron, S.H. Bennekou, Towards a regulatory use of alternative developmental neurotoxicity testing (DNT), *Toxicol Appl Pharmacol* 354 (2018) 19-23. [10.1016/j.taap.2018.02.002](https://doi.org/10.1016/j.taap.2018.02.002)
- [19] OECD, 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 & Assessment 256 (2016).
- [20] OECD, Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment, in: OECD (Ed.) Series on Testing and Assessment 255, 2016.
- [21] L. Meigs, L. Smirnova, C. Rovida, M. Leist, T. Hartung, Animal testing and its alternatives - the most important omics is economics, *ALTEX* 35(3) (2018) 275-305. [10.14573/altex.1807041](https://doi.org/10.14573/altex.1807041)
- [22] S.L. Makris, K. Raffaele, S. Allen, W.J. Bowers, U. Hass, E. Alleva, G. Calamandrei, L. Sheets, P. Amcoff, N. Delrue, K.M. Crofton, A retrospective performance assessment of the developmental neurotoxicity study in support of OECD test guideline 426, *Environ Health Perspect* 117(1) (2009) 17-25. [10.1289/ehp.11447](https://doi.org/10.1289/ehp.11447)
- [23] W.R. Mundy, S. Padilla, J.M. Breier, K.M. Crofton, M.E. Gilbert, D.W. Herr, K.F. Jensen, N.M. Radio, K.C. Raffaele, K. Schumacher, T.J. Shafer, J. Cowden, Expanding the test set: Chemicals with potential to disrupt mammalian brain development, *Neurotoxicol Teratol* 52(Pt A) (2015) 25-35. [10.1016/j.ntt.2015.10.001](https://doi.org/10.1016/j.ntt.2015.10.001)
- [24] W.M.S. Russell, R.L. Burch, *The Principles of Humane Experimental Technique*, 1959.
- [25] V. Moser, F. Bailey, W. Bowers, K. Raffaele, K. Croton, M. Gilbert, Developmental Neurotoxicity Study Guidance Document, North American Free trade Agreement (NAFTA) Technical Working Group on Pesticides, 2016.
- [26] K.C. Raffaele, J.E. Fisher, S. Hancock, K. Hazelden, S.K. Sobrian, Determining normal variability in a developmental neurotoxicity test: a report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints, *Neurotoxicol Teratol* 30(4) (2008) 288-325. [10.1016/j.ntt.2007.12.009](https://doi.org/10.1016/j.ntt.2007.12.009)
- [27] J. Buelke-Sam, C.A. Kimmel, J. Adams, C.J. Nelson, C.V. Vorhees, D.C. Wright, V. St Omer, B.A. Korol, R.E. Butcher, M.A. Geyer, et al., Collaborative Behavioral Teratology Study: results, *Neurobehav Toxicol Teratol* 7(6) (1985) 591-624. <https://www.ncbi.nlm.nih.gov/pubmed/3835454>
- [28] A.W. Hayes, A.D. Dayan, W.C. Hall, R.L. Kodell, G.M. Williams, W.D. Waddell, R.S. Slesinski, C.L. Kruger, A review of mammalian carcinogenicity study design and potential effects of alternate test procedures on the safety evaluation of food ingredients, *Regul Toxicol Pharmacol* 60(1 Suppl) (2011) S1-34. [10.1016/j.yrtph.2010.10.002](https://doi.org/10.1016/j.yrtph.2010.10.002)
- [29] K. Schneider, M. Schwarz, I. Burkholder, A. Kopp-Schneider, L. Edler, A. Kinsner-Ovaskainen, T. Hartung, S. Hoffmann, "ToxRTool", a new tool to assess the reliability of toxicological data, *Toxicol Lett* 189(2) (2009) 138-44. [10.1016/j.toxlet.2009.05.013](https://doi.org/10.1016/j.toxlet.2009.05.013)
- [30] C. Rudén, The use and evaluation of primary data in 29 trichloroethylene carcinogen risk assessments, *Regul Toxicol Pharmacol* 34(1) (2001) 3-16. [10.1006/rtph.2001.1482](https://doi.org/10.1006/rtph.2001.1482)
- [31] European Commission, Thresholds for Endocrine Disrupters and Related Uncertainties. Report of the Endocrine Disrupters Expert Advisory Group, European Commission, 2013.
- [32] US-EPA, State of the Science Evaluation: Nonmonotonic Dose Responses as They Apply to Estrogen, Androgen, and Thyroid Pathways and EPA Testing and Assessment Procedures, State of the Science: Nonmonotonic Dose Responses V7, US, EPA, 2013.
- [33] R. Tsuji, K.M. Crofton, Developmental neurotoxicity guideline study: issues with methodology, evaluation and regulation, *Congenit Anom (Kyoto)* 52(3) (2012) 122-8. [10.1111/j.1741-4520.2012.00374.x](https://doi.org/10.1111/j.1741-4520.2012.00374.x)
- [34] S.K. Bopp, A. Kienzler, A.N. Richarz, S.C. van der Linden, A. Paini, N. Parissis, A.P. Worth, Regulatory assessment and risk management of chemical mixtures: challenges and ways forward, *Crit Rev Toxicol* 49(2) (2019) 174-189. [10.1080/10408444.2019.1579169](https://doi.org/10.1080/10408444.2019.1579169)

- [35] F. Pistollato, E.M. De Gyves, D. Carpi, S.K. Bopp, C. Nunes, A. Worth, A. Bal-Price, Assessment of developmental neurotoxicity induced by chemical mixtures using an adverse outcome pathway concept, *Environmental Health* 19(1) (2020). 10.1186/s12940-020-00578-x
- [36] OECD, Retrospective performance assessment of the test guideline 426 on developmental neurotoxicity., *Series on testing and assessment* 89 (2008).
- [37] P.J. Catalano, K.L. McDaniel, V.C. Moser, The IPCS Collaborative Study on Neurobehavioral Screening Methods: VI. Agreement and reliability of the data, *Neurotoxicology* 18(4) (1997) 1057-64. <https://www.ncbi.nlm.nih.gov/pubmed/9457734>
- [38] R.C. MacPhail, H.A. Tilson, V.C. Moser, G.C. Becking, V. Cuomo, E. Frantik, B.M. Kulig, G. Winneke, The IPCS Collaborative Study on Neurobehavioral Screening. I. Background and genesis, *Neurotoxicology* 18(4) (1997) 925-8. <https://www.ncbi.nlm.nih.gov/pubmed/9457729>
- [39] C. Dumont, J. Barroso, I. Matys, A. Worth, S. Casati, Analysis of the Local Lymph Node Assay (LLNA) variability for assessing the prediction of skin sensitisation potential and potency of chemicals with non-animal approaches, *Toxicol In Vitro* 34 (2016) 220-8. 10.1016/j.tiv.2016.04.008
- [40] S. Hoffmann, LLNA variability: An essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies, *ALTEX* 32(4) (2015) 379-83. 10.14573/altex.1505051
- [41] D.W. Roberts, A.M. Api, A.O. Aptula, Chemical applicability domain of the Local Lymph Node Assay (LLNA) for skin sensitisation potency. Part 2. The biological variability of the murine Local Lymph Node Assay (LLNA) for skin sensitisation, *Regul Toxicol Pharmacol* 80 (2016) 255-9. 10.1016/j.yrtph.2016.07.013
- [42] E. Gottmann, S. Kramer, B. Pfahringer, C. Helma, Data quality in predictive toxicology: reproducibility of rodent carcinogenicity experiments, *Environ Health Perspect* 109(5) (2001) 509-14. 10.1289/ehp.01109509
- [43] OECD, Developmental Neurotoxicity Study, *OECD Guidelines for the Testing of Chemicals* 426 (2007).
- [44] OECD, Extended One Generation Reproductive Toxicity Study, *OECD Guidelines for the Testing of Chemicals* 443 (2018).
- [45] UK-HSE, Investigation of the state of the art on identification of appropriate reference points for the derivation of health-based guidance values (ADI, AOEL and AAOEL) for pesticides and on the derivation of uncertainty factors to be used in human risk assessment, *Chemicals Regulation Directorate, Health & Safety Executive, UK, 2013, p. 169pp.*
- [46] E.Z. Francis, C.A. Kimmel, D.C. Rees, Workshop on the qualitative and quantitative comparability of human and animal developmental neurotoxicity: summary and implications, *Neurotoxicol Teratol* 12(3) (1990) 285-92. 10.1016/0892-0362(90)90101-h
- [47] M. Aschner, S. Ceccatelli, M. Daneshian, E. Fritsche, N. Hasiwa, T. Hartung, H.T. Hogberg, M. Leist, A. Li, W.R. Mundi, S. Padilla, A.H. Piersma, A. Bal-Price, A. Seiler, R.H. Westerink, B. Zimmer, P.J. Lein, Reference compounds for alternative test methods to indicate developmental neurotoxicity (DNT) potential of chemicals: example lists and criteria for their selection and use, *ALTEX* 34(1) (2017) 49-74. 10.14573/altex.1604201
- [48] D.C. Dorman, S.L. Allen, J.Z. Byczkowski, L. Claudio, J.E. Fisher, J.W. Fisher, G.J. Harry, A.A. Li, S.L. Makris, S. Padilla, L.G. Sultatos, B.E. Mileson, Methods to identify and characterize developmental neurotoxicity for human health risk assessment. III: pharmacokinetic and pharmacodynamic considerations, *Environ Health Perspect* 109 Suppl 1 (2001) 101-11. 10.1289/ehp.01109s1101
- [49] W. Kaufmann, Current status of developmental neurotoxicity: an industry perspective, *Toxicol Lett* 140-141 (2003) 161-9. 10.1016/s0378-4274(02)00503-9
- [50] M. Paparella, M. Daneshian, R. Hornek-Gausterer, M. Kinzl, I. Mauritz, S. Mühlegger, Uncertainty of testing methods--what do we (want to) know?, *ALTEX* 30(2) (2013) 131-44. 10.14573/altex.2013.2.131
- [51] M.E. McDonald, J.F. Paul, Timing of increased autistic disorder cumulative incidence, *Environ Sci Technol* 44(6) (2010) 2112-8. 10.1021/es902057k
- [52] M.J. Carey, Comment on "timing of increased autistic disorder cumulative incidence", *Environ Sci Technol* 46(3) (2012) 1948-9; author reply 1950-1. 10.1021/es2037965
- [53] T.E. Bakken, J.A. Miller, S.-L. Ding, S.M. Sunkin, K.A. Smith, L. Ng, A. Szafer, R.A. Dalley, J.J. Royall, T. Lemon, S. Shapouri, K. Aiona, J. Arnold, J.L. Bennett, D. Bertagnolli, K. Bickley, A. Boe, K. Brouner, S.

- Butler, E. Byrnes, S. Caldejon, A. Carey, S. Cate, M. Chapin, J. Chen, N. Dee, T. Desta, T.A. Dolbeare, N. Dotson, A. Ebbert, E. Fulfs, G. Gee, T.L. Gilbert, J. Goldy, L. Gourley, B. Gregor, G. Gu, J. Hall, Z. Haradon, D.R. Haynor, N. Hejazinia, A. Hoerder-Suabedissen, R. Howard, J. Jochim, M. Kinnunen, A. Kriedberg, C.L. Kuan, C. Lau, C.-K. Lee, F. Lee, L. Luong, N. Mastan, R. May, J. Melchor, N. Mosqueda, E. Mott, K. Ngo, J. Nyhus, A. Oldre, E. Olson, J. Parente, P.D. Parker, S. Parry, J. Pendergraft, L. Potekhina, M. Reding, Z.L. Riley, T. Roberts, B. Rogers, K. Roll, D. Rosen, D. Sandman, M. Sarreal, N. Shapovalova, S. Shi, N. Sjoquist, A.J. Sodt, R. Townsend, L. Velasquez, U. Wagley, W.B. Wakeman, C. White, C. Bennett, J. Wu, R. Young, B.L. Youngstrom, P. Wohnoutka, R.A. Gibbs, J. Rogers, J.G. Hohmann, M.J. Hawrylycz, R.F. Hevner, Z. Molnár, J.W. Phillips, C. Dang, A.R. Jones, D.G. Amaral, A. Bernard, E.S. Lein, A comprehensive transcriptional map of primate brain development, *Nature* 535(7612) (2016) 367-375. 10.1038/nature18637
- [54] Juan, Edward, S.J. Caddick, P. Churchland, G. Feng, Gregg, K.-F. Lee, David, Cory, Jude, S. Mitalipov, Alysson, J. H. Okano, John, D.L. Ringach, Terrence, Afonso, Peter, J. Wu, F. Zhang, Brains, Genes, and Primates, *Neuron* 86(3) (2015) 617-631. 10.1016/j.neuron.2015.03.021
- [55] K. Dach, F. Bendt, U. Huebenthal, S. Giersiefer, P.J. Lein, H. Heuer, E. Fritsche, BDE-99 impairs differentiation of human and mouse NPCs into the oligodendroglial lineage by species-specific modes of action, *Sci Rep* 7 (2017) 44861. 10.1038/srep44861
- [56] J. Baumann, K. Gassmann, S. Masjosthusmann, D. DeBoer, F. Bendt, S. Giersiefer, E. Fritsche, Comparative human and rat neurospheres reveal species differences in chemical effects on neurodevelopmental key events, *Arch Toxicol* 90(6) (2016) 1415-27. 10.1007/s00204-015-1568-8
- [57] J.A. Harrill, T.M. Freudenrich, B.L. Robinette, W.R. Mundy, Comparative sensitivity of human and rat neural cultures to chemical-induced inhibition of neurite outgrowth, *Toxicol Appl Pharmacol* 256(3) (2011) 268-80. 10.1016/j.taap.2011.02.013
- [58] K. Gassmann, J. Abel, H. Bothe, T. Haarmann-Stemann, H.F. Merk, K.N. Quasthoff, T.D. Rockel, T. Schreiber, E. Fritsche, Species-specific differential AhR expression protects human neural progenitor cells against developmental neurotoxicity of PAHs, *Environ Health Perspect* 118(11) (2010) 1571-7. 10.1289/ehp.0901545
- [59] A. De Felice, L. Ricceri, A. Venerosi, F. Chiarotti, G. Calamandrei, Multifactorial Origin of Neurodevelopmental Disorders: Approaches to Understanding Complex Etiologies, *Toxics* 3(1) (2015) 89-129. 10.3390/toxics3010089
- [60] E. Fritsche, H. Alm, J. Baumann, L. Geerts, H. Håkansson, S. Masjosthusmann, H. Witters, Literature review on in vitro and alternative Developmental Neurotoxicity (DNT) testing methods, EFSA Supporting publication 2015:EN-778 (2015) 1-186. 10.2903/sp.efsa.2015.EN-778
- [61] A. Maertens, N. Anastas, P.J. Spencer, M. Stephens, A. Goldberg, T. Hartung, Green toxicology, *ALTEX* 31(3) (2014) 243-9. 10.14573/altex.1406181
- [62] C. Sharbaugh, S.M. Viet, A. Fraser, S.B. McMaster, Comparable measures of cognitive function in human infants and laboratory animals to identify environmental health risks to children, *Environ Health Perspect* 111(13) (2003) 1630-9. 10.1289/ehp.6205
- [63] M. Sarter, Animal cognition: defining the issues, *Neurosci Biobehav Rev* 28(7) (2004) 645-50. 10.1016/j.neubiorev.2004.09.005
- [64] Y. Chudasama, J.L. Muir, A behavioural analysis of the delayed non-matching to position task: the effects of scopolamine, lesions of the fornix and of the prelimbic region on mediating behaviours by rats, *Psychopharmacology (Berl)* 134(1) (1997) 73-82. 10.1007/s002130050427
- [65] C.V. Vorhees, J.N. Spowles, S.L. Regan, M.T. Williams, A better approach to in vivo developmental neurotoxicity assessment: Alignment of rodent testing with effects seen in children after neurotoxic exposures, *Toxicol Appl Pharmacol* 354 (2018) 176-190. 10.1016/j.taap.2018.03.012
- [66] K. von Stackelberg, E. Guzy, T. Chu, B. Claus Henn, Exposure to Mixtures of Metals and Neurodevelopmental Outcomes: A Multidisciplinary Review Using an Adverse Outcome Pathway Framework, *Risk Anal* 35(6) (2015) 971-1016. 10.1111/risa.12425
- [67] A. Bal-Price, M.E.B. Meek, Adverse outcome pathways: Application to enhance mechanistic understanding of neurotoxicity, *Pharmacol Ther* 179 (2017) 84-95. 10.1016/j.pharmthera.2017.05.006