Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Toxicology

Regulatory use and acceptance of alternative methods for chemical hazard identification

Patience Browne, Nathalie Delrue and Anne Gourmelon

Abstract

Despite the expansion of methods and technologies available for evaluating the safety of environmental chemicals, the uptake of new approaches and acceptance of alternative data in regulatory contexts have been relatively slow. This may be due to real limitations of alternative methods, as well as the perception that 'traditional', animal-based toxicological methods are more protective of human health, although recent meta-analyses of large data sets indicate the contrary in some cases. Animal data often are weighted more heavily in chemical hazard identifications than results derived from alternative methods, particularly when alternative data are negative. We identified several science-based limitations in alternative methods and propose approaches to reduce the limitations and increase confidence in (particularly negative) results. We also suggest that the limitations of animal data should be clearly communicated to avoid holding nonanimal alternatives to unrealistic performance standards and predictors of human health. Until the chemical industry can be confident that both positive and negative alternative data will be accepted and regulators can be confident that alternative data are good predictors of the toxicological response, animal tests will continue to be used (where not prohibited) as methods to unequivocally satisfy regulatory data requirements.

Addresses

Organisation for Economic and Cooperative Development, Environment, Health and Safety Division, Environmental Directorate, 2, Rue Andre-Pascal, 75775 Paris Cedex 16, France

Corresponding author: Browne, Patience (patience.browne@oecd.org)

Current Opinion in Toxicology 2019, **15**:18–25

This review comes from a themed issue on **Risk assessment in toxicology**

Available online 16 February 2019

For a complete overview see the Issue and the Editorial

https://doi.org/10.1016/j.cotox.2019.02.003

Keywords

Alternative methods, Hazard identification, Chemical regulation, Predictive performance.

Introduction

Technologies developed over the past two decades have produced extraordinary innovations in the methods available for toxicological testing, including a variety of alternative approaches[1] to traditional animal-based chemical safety evaluation. Many *in vivo* toxicity tests were developed in the first half of the last century and rely on relatively insensitive measures such as organ weight changes and general endpoints such as decreased growth, survival, or reproduction measured in a few animals exposed to widely spaced doses of a test chemical. Methods are now available to measure cellular, biochemical, and -omics endpoints considerably 'upstream' of these effects and at greatly increased sensitivity. Alternative methods can be used to prioritize chemicals for further testing, as preliminary screens, as part of testing strategies, or may predict the outcome of animal testing with sufficient accuracy to replace *in vivo* tests. Furthermore, the strength of alternative test systems is that they can be deployed rapidly, applied to large numbers of chemicals, and completed quickly and at relatively manageable cost. Despite the potential increase in chemical throughput, infiltration of alternative methods in regulatory decision-making has failed to keep pace with emerging science or to make substantial progress in evaluating chemical backlog awaiting safety decisions e.g. [1,2].

Much of the innovation in chemical safety screening has been driven by the strategic vision detailed in the National Academies of Science Toxicity Testing in the 21st Century [3], which helped foster the swift expansion of technologies that are increasingly available for evaluating chemical hazards, and in turn, has led to changes in regulatory data requirements and acceptance. For example, the European Union (EU) imposed a ban on animal testing of cosmetics ingredients in 2013, and a recent European Chemicals Agency (ECHA) report states that animal testing should be used only as a last resort [4]. Other countries, such as India, have adopted similar regulations. The US Environmental Protection Agency (EPA) has committed to finding suitable alternatives for animal testing of pesticides [5] and industrial chemicals [6], and the US Food and Drug Administration (FDA) recently published a strategic vision

[1] For the purpose of this discussion "alternative approaches", include *in silico* modeling tools, *in chemico* and *in vitro* tests, and mining and analyses of existing databases to satisfy testing requirements and determine additional testing needs.

describing a roadmap for increasing reliance on alternative approaches [7]. Conversely, some regulations require knowledge of the chemical toxicity mechanism in association with the demonstration of adversity in an intact organism (e.g. regulation of endocrine disrupting chemicals in the US and EU).

The Organisation for Economic Cooperation and Development (OECD) assists countries in validating and harmonizing tools for international chemical safety evaluation. Data generated using an OECD Test Guideline following the principles of Good Laboratory Practice are subject to the agreement on the Mutual Acceptance of Data, and thus, results must be accepted by all OECD member and partner countries that have a regulatory requirement for that endpoint. In addition to reducing animal use from duplicative testing, the OECD is committed to finding alternatives to animal testing when the reliability and relevance of the alternative can be adequately demonstrated. To this end, the OECD has developed a number of alternative tools for chemical safety evaluation, including test guidelines and associated guidance documents to help determine testing needs, and the OECD QSAR Toolbox for *in silico* predictions of chemical activity (http://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm). In some cases, *in vitro* methods are accepted as alternatives for *in vivo* guideline methods, including four of the most frequently required regulatory endpoints (skin irritation/corrosion, eye irritation, skin sensitization, and mutagenicity). More recently, the OECD has developed testing strategies, sequential or integrated deployment of alternative methods that can reduce or eliminate the need for animal testing.

Although alternative methods are beginning to be used to evaluate chemical safety, there is reticence to accept alternative data for regulatory decision-making, particularly when results are negative e.g. [8]. The regulatory purview is to protect human health and the environment and thus must be conservative regarding the data accepted for this purpose. When evaluating the suitability of alternative data for regulatory purposes, it is important to discriminate technical/scientifically based limitations in alternative methods from an anachronistic philosophy that animal data are 'better' and more informative of human health hazard. Herein, we discuss limitations in *in vivo* animal data, discuss potential biases in weighting *in vivo* and *in vitro* data in chemical safety evaluations, and provide suggestions for improving confidence in alternative test method data for regulatory decisions.

Recognition of limitations of *in vivo* data

In addition to animal welfare, chemical throughput, and cost considerations, several recent publications question rodent toxicity data as 1) a 'gold standard' measure for human health and 2) appropriate reference data for *in vitro* assay development for some endpoints [9–12]. Physiological processes such as gametogenesis, gestation, and early neonatal development are recognized windows of susceptibility to toxicants, although the marked differences in critical life stages between humans and rodents challenge the predictivity of the rodent experimental model [12]. Furthermore, chemicals tested in rodents may be positive (or negative) by mechanisms that are not relevant to humans [13]. Recently, Cohen [11] concluded the two-year rodent bioassay had an unacceptably high false-positive rate and thus is not useful for identifying human mutagens or genotoxins. Several analyses of guideline animal studies indicated that even for simple, well-understood toxicological processes, rodents may be poor predictors of the human response. The rodent local lymph node assay (LLNA) had a 14–20% false-positive rate when compared to human skin sensitization responses [14]. More deleterious, several high profile examples of drugs that were tested extensively in animals with no adverse response caused serious effects in humans [15].

Analyses of guideline animal studies indicate that results of chemicals tested more than once had limited reproducibility [10,16,17]. An analysis of chemicals in the REACH database that were tested more than once following the same *in vivo* guideline indicated positive/negative results were reproducible for 72–94% of the chemicals [17]. Differences in animal strain and husbandry are contributing factors, although (in these analyses) the differences are allowable in guideline methods, and therefore represent a contribution to uncertainty in regulatory decisions [9]. The reproducibility of chemical safety data is dependent on the chemical itself; highly toxic and nontoxic chemicals give more reproducible results than moderate and weakly potent chemicals [10,17]. Specificity (i.e. conditional probability that a chemical was initially negative *in vivo* and negative upon retesting) was higher (82–100%) because of the high proportion of environmental chemicals that are negative for any particular toxicological effect [17]. Conversely, the sensitivity of the animal data (conditional probability that a chemical is initially positive *in vivo* and positive upon retesting) ranged from 50 to 87% [17].

To help dispel the idea that *in vivo* data are 'superior' and more protective of human health than alternative methods, uncertainty and limitation of the animal studies should be better documented. Additional systematic metadata analyses may further indicate where *in vivo* data are actually less protective of human health than other approaches.

Weighting of *in vivo* and *in vitro* results

In general, regulators seem to be more comfortable considering positive alternative data than negative results in chemical safety evaluations or to satisfy an *in vivo* testing requirement. A review of US EPA's acceptance of existing data submitted to satisfy Tier 1 endocrine screening data requirements suggests a preferential acceptance of positive data. Of the 67 chemicals that were included in endocrine disruptor screening program (EDSP) List 1, registrants for 47 chemicals requested testing waivers based on data considered equivalent to at least some of the 11 assays in the Tier 1 screening battery. The US EPA granted waivers based on existing positive study results much more often than for negative results (64 studies versus 25 studies) [8].

Recent publications identified mechanistic characteristics of carcinogens as a way to organize and evaluate carcinogenic hazard [18,19]. Because mechanistic assays are not available for all the key characteristics of carcinogens, conclusions cannot be determined from mechanistic data alone [20,18]. However, positive mechanistic data may upgrade chemicals with inadequate/limited information from human and animal data to be reconsidered as having limited/sufficient data and may alter the classification of chemicals to 'carcinogenic in humans' (group 1) and 'probably carcinogenic in humans' (group 2A) [21]. In general, downgrades of chemicals from 'possibly carcinogenic in humans' (group 2B) to 'not classifiable' (group 3) based on mechanistic data occurred only when the mechanism was determined not to be relevant in humans [20]. In a recent meta-analysis of the International Agency for Research on Cancer (IARC) evaluation of the potential of 34 agents to cause cancer in humans, the inclusion of mechanistic data resulted in a change from group 2B to group 2A for 2 of 34 (6%) agents [21]. In fact, more mechanistic data were available for well-studied agents with accompanying human and animal data [21], so in these instances, mechanistic data may have contributed less to the overall hazard identification than in circumstances where human data were considered 'limited' or 'inadequate'. Given the recognized high false-positive rate of the rodents carcinogenicity assay, considering negative outcomes of *in vitro* assays measuring key characteristics of carcinogens may actually improve predictions of human carcinogens in cases where limited human data are available.

In some circumstances, *in vitro* test methods may have clearly defined criteria under which negative test results cannot reliably indicate the absence of a potential effect. For example, for Test Guidelines 442D/E, all positive results are accepted; however, negative results for chemicals with octanol–water partition coefficients ($\log K_{ow}$) > 3.5 are outside of the applicability domain of the test [22,23].

There are a few examples where negative *in vitro* data are accepted, in principle, as sufficient evidence of absence of effects. This is the case with certain *in vitro* methods for skin or eye irritation using reconstructed three-dimensional (3D) tissues or organotypic assays, where negative outcomes of the assay are interpreted as absence of irritation potential of the chemical and no further testing is needed to confirm that conclusion [24,25]. Requirements for genotoxicity data for biocides require three *in vitro* assays [26] and industrial chemicals require four *in vitro* assays; in both cases, if all *in vitro* data are negative, no animal data are required. Interestingly, pesticide chemical registration requires three *in vitro* assays for genotoxicity, and even when all *in vitro* data are negative, one *in vivo* study is required for confirmation (EU Regulation 283/2013). Nevertheless, in practice, regulators having access to *in vivo* data will tend to credit animal data more than *in vitro* data, especially when there is conflicting information where *in vitro* data are negative and animal data are positive, by organizing data in a tiered approach where animal data are considered superior [27]. Furthermore, if there is substantial chemical space for which no prediction can be made or there is a 'grey zone' under which uncertainty is too large, industry will not incur the costs of using alternative tests and, instead, go directly to the animal study that will be accepted, if not specifically prohibited by regulations. Thus, to incentivize use, alternative approaches should be accompanied by clear interpretation procedures and a description of limitations for both users of the approaches and recipients of the data.

Recognition of limitations of *in vitro* data

Negative data resulting from alternative methods are often considered with caveats, following the rationale that nonanimal test systems cannot account for complex processes that occur in humans and therefore should not be accepted in the absence of an *in vivo* confirmation (of a lack of an effect). Although recent evaluations spotlight some limitations of the rodent as a model for human toxic responses, *in vivo* tests can account for all mechanisms by which a chemical may exert an adverse effect in the exposed animal. In absence of evidence to the contrary, chemical effects in rodent models are assumed to be relevant to humans. On the other hand, *in vitro* systems are inherently reductive and cannot fully account for xenobiotic metabolism and toxicokinetic effects that influence toxicity in whole organisms, nor can they capture the influence of other factors toxicologists are just beginning to grapple with, such as the contribution of microbial metabolism to chemical toxicity [28]. Furthermore, all *in vitro* methods have inherent limitations defined by the physical–chemical properties of the test chemical which affect its interaction with the test system (e.g. solubility, volatility, molecular weight) or the detection system (e.g.

cytotoxicity, protein denaturing, fluorescent compounds) [29,30]. While the applicability domain of *in vivo* assays is assumed to be the chemical universe, the applicability domain of the *in vitro* assay is defined by reference chemicals used to validate the assay [31].

Considerations to improve regulatory acceptance of alternative data

For users to produce and accept alternative data in lieu of *in vivo* results, they must be able to trust positive and negative outcomes. Therefore, a critical step for improving regulatory acceptance of alternative data is improving confidence in negative results, and to achieve this, it is important to separate science-based limitations of methods from an assumption that rodent data are a superior indicator of human health effects. In the following paragraphs, we suggest several areas for improving alternative method predictions, focused largely on improving confidence in negative resulting data.

Reference chemical selection

Part of validating a test method is determining the ability of the method to identify positive and negative reference chemicals. Historically, *in vivo* validation efforts included mostly positive reference chemicals and only one or two negative chemicals. Even with current high-throughput methods able to evaluate performance against much larger numbers of chemicals, reference chemicals are usually proportionally $\geq$60% positive chemicals (http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm). Yet, an evaluation of the 9000 chemicals included in the REACH database indicated fewer than 20% of chemicals were positive for any specific type of toxicity [32]. Guideline methods used for chemical regulation are optimized to reduce false negatives (i.e. high sensitivity), but as a consequence of optimized sensitivity, false-positive rates can become unacceptably high. Identification of negative reference chemicals is challenging [33], but until reference sets include positive/negative chemicals at proportions resembling the expected occurrence of the environmental toxic hazards, toxicological methods are likely to overpredict hazards. As mentioned previously, potently toxic chemicals and nontoxic chemicals are correctly identified at a higher rate than weakly potent chemicals. Selection of weakly positive chemicals that are expected to have environmentally relevant potencies will help to improve predictions of chemicals that are neither negative nor highly toxic. In some cases, high false-positive rates may be acceptable or preferable to regulators, but in these cases, this limitation in the assay should be clearly communicated.

Assay interference

Some chemicals are capable of reacting with assay reagents or biological molecules, resulting in nonspecific, spurious results [30,34]. This is a particular challenge in loss-of-function assays (e.g. enzyme inhibition, nuclear receptor antagonism), where it is necessary to distinguish 'true' activity from a decrease in signal due to cytotoxicity. However, more subtle effects such as promiscuous or 'pan-reactive' compounds can be difficult to identify [34,29,30]. Suites of assays that measure chemical interaction with a specific mechanism using different technologies can be helpful for elucidating compounds that interfere system readouts [35], although redundant assays are not available for all molecular targets. As an alternative, current efforts are underway to evaluate large chemicals libraries screened for interactions with, in some cases, hundreds of chemical targets to identify a set of assay interference reference chemicals that could be used to classify limitations of assays and determine chemistries that may produce false-positive/false-negative results [36,34,29].

Alternative methods to better recapitulate *in vivo* biology

A frequent criticism of *in vitro* test systems is their limited xenobiotic metabolic capabilities that may not adequately capture biological modification of parent compounds to potentially more or less potent metabolites [37,38]. Some cell-based test systems have native [39] or specifically engineered [40,41] xenobiotic metabolic capabilities. And, while there have been very few *in vitro* experiments that evaluate the bioactivity in parent chemical and their metabolites, the few data available indicate that some chemicals were biotransformed to more potent metabolites, but there were no examples identified of completely negative parent chemicals that were biotransformed to potent metabolites [42,43]. Ongoing efforts to mimic chemical biotransformation *in vitro* (and *in silico*) will likely improve both *in vitro* potency estimates and confidence in resulting data. New 3D culture systems more closely resemble human physiology than two-dimensional monolayer cultures, although recent publications report differences in the xenobiotic metabolism of HepaRG cells in different 3D models [44] and highlight the importance of characterizing xenobiotic metabolic capabilities of the *in vitro* system [45]. Coincubation of the parent compound with S9 fractions or other introduced simulators of hepatic xenobiotic metabolism are near-term approaches [37] until more sophisticated *in vitro* test systems become standardized and validated for regulatory use. Using a set of relevant compounds to characterize xenobiotic enzyme activity [46] can help address the caveat that *in vitro* systems are missing potent metabolites of parent chemicals.

In addition to xenobiotic metabolism, an understanding of toxicokinetic processes that occur in organisms is needed to extrapolate active concentrations of test chemical in *in vitro* systems to an *in vivo* dose. A variety of *in silico* and *in vitro* toxicokinetic methods exist for

translating active concentration to human tissues doses, and efforts are underway to systemically characterize and compare these approaches [46] and create databases to house *in vitro*/*in vivo* data to improve models [47,48]. Currently, *in vitro* data are often interpreted as positive/negative or binned into potency categories, rather than using the full range of quantitative resulting information available. With physiologically based kinetic models, *in vitro* data can be correlated to *in vivo* responses (e.g. dose—response, point of departure, threshold limits) [49,50].

Alternative methods are being continually improved to better mimic processes that occur in organisms, but the relevance of *in vitro* effects to humans is often questioned. Where available, human cell-based alternative methods are inherently relevant for human safety evaluation, and for some endpoints, these alternative methods may better predict human effects that rodents e.g. [51].

Transparency regarding limitations in animal data

When it was adopted as a toxicological model, the rodent was assumed to be relevant to human health. If validated at all, the reliability and relevance of rodent assays were determined from a handful of chemicals, generally without absorption/distribution/metabolism/excretion (ADME) information or internal exposure. Reference chemicals were often potent pharmaceuticals and toxicants known to produce the frank toxicological effect, and validation of the animal model did not typically include the dynamic range of *in vivo* responses to weaker positive chemicals more relevant for human environmental exposure. While long-term *in vivo* studies measure unequivocal effects (e.g. carcinogenicity, development, survival, longevity, reproduction), it can be difficult to distinguish adaptive responses from persistent, consequential effects in short-term studies that examine organ-level endpoints.

Except in the few cases where high-quality human data are available, results of animal tests are used to interrogate the performance of alternative methods. There are now a variety of examples illustrating endpoints for which rodent responses are not relevant to humans [52—54], and the limitations of animal tests as a reliable and relevant model for humans must be acknowledged and considered in the evaluation of any alternative for an existing *in vivo* test. To validate an *in vitro* alternative for a rodent method, a much larger number of chemicals tested at concentrations ranging over several orders of magnitude are now required to characterize the chemical domain for which the assay can be used, and the performance is expected to be equal to or exceed the performance of the existing test methods [55]. Metadata analyses of animal toxicology data accumulated over the past five decades indicate the overall reproducibility of *in vivo* guideline results is about 75%, and therefore, the predictivity of any alternative method for the animal test cannot be expected to exceed that level. In an example of a toxicological endpoint with substantial *in vitro*, *in vivo*, and human reference data, the *in vitro* methods outperform *in vivo* data for predicting human responses [56]. But, rather than contributing to a sense of higher standards of 'validation', this more rigorous validation process seems to have fostered a sense that resulting data should be considered with greater misgivings.

Testing strategies

Until recently, single alternative methods were evaluated as a potential replacement for an exiting *in vivo* test. In many cases, animal tests include a variety of endpoints and biological complexity that cannot be adequately simulated in a single *in vitro* test. Furthermore, this one-for-one consideration of *in vivo* test replacement fails to leverage the strength of alternative methods that are capable of rapidly screening a library of chemicals for a variety of endpoints and may use a battery of methods that overcome the limitations of any single *in vitro* assay. For example, the reported limitations of sensitivity in *in vivo* developmental neurotoxicity endpoints can be improved by using a battery of *in vitro* assays and may provide a better mechanistic understanding of the disease process [57], even for complex endpoints such as cancer [11,58].

In vitro test methods may provide a mechanistic understanding of the induced toxicity, but it can be difficult to link a mechanism to an *in vivo* effect. These gaps in the understanding of biological relationships between *in vitro* mechanism and *in vivo* adverse effects have been a constraint to using alternative approaches for regulatory decision-making. To overcome this challenge, the concepts of adverse outcome pathways (AOPs) were introduced to relate molecular events to downstream effects on cells, organs, tissues, and individuals. AOPs were introduced as a framework for integrating diverse data measured at different levels of biological organization and linking mechanistic information to endpoints used for regulatory purposes [59,60]. Testing strategies that use multiple sequential or integrated alternative methods can be developed around AOP frameworks [61] and can be applied to a regulatory context [62,63].

Conclusions

The current examples of endpoints with robust alternatives accepted for regulatory decisions in lieu of *in vivo* data tend to be for endpoints where toxicity is triggered by a finite number of well-understood mechanisms (e.g. skin sensitization [63,64]) or where xenobiotic metabolism does not have a substantial influence on the effect (e.g. phototoxicity [65]). While these examples may be the exception to the rule, identifying the scientific and nonscientific challenges to regulatory

acceptance of alternative data for these 'simple cases' can help reduce limitations of alternative methods, address obstacles relevant to more complex endpoints, and foster overall confidence for use of alternative data in regulatory contexts.

Herein, we highlight several science-based considerations to more thoroughly and transparently characterize alternative methods, which should help improve the confidence in resulting data, and particularly, negative results. Reference chemicals that include a balance of positive/negative chemicals with an emphasis on weakly positive environmentally relevant chemicals will likely improve the specificity of assays. Identifying reference chemicals that interact with assay through nonspecific mechanisms and characterizing their effect on the test system, along with positive and negative reference chemicals, define the domain of applicability of the alternative method. A variety of *in vitro* and *in silico* approaches are being developed to better account for ADME influences on the chemical hazard. These range from coincubating test chemical with S9 fractions and computational approaches to testing strategies and 3D cell and organoid models. These innovative approaches take advantage of the strengths of alternative approaches and use multiple methods in combination to overcome limitations of a single method. Finally, the limitations of animal data can be more transparently captured and communicated to help dispel the (mis) conception that in every case, animal data provide a better prediction for the human response.

As global animal welfare and testing needs for chemical regulations become increasingly disparate, the restrictions on animal testing in some areas, contrasted with the obstacles for accepting alternative data for regulatory decisions in others, run the risk of interfering with the mutual acceptance of data and thus increasing animal testing, globally. While broader regulatory acceptance of alternative approaches is beginning, there is currently more willingness to accept positive results (indicating the test chemical represents a potential hazard) than negative results. Until the uncertainties and limitations of *in vivo* standards and alternative data are clearly communicated and regulators can be confident in both positive and negative outcomes of alternative methods, alternatives for animal testing will not be 'true' replacements for *in vivo* data.

Disclaimer

The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the OECD or its member countries.

Conflict of interest statement

Nothing declared.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. NTP: *National Toxicology Program. Testing information*. 2017 [Online]. Available: https://ntp.niehs.nih.gov/testing/index.html.
2. ECHA: *European Chemicals Agency. Registered substances database*. 2018 [Online]. Available: https://echa.europa.eu/information-on-chemicals/registered-substances.
3. *Toxicity testing in the 21st century*. Washington, D.C.: National Academies Press; 2007.
4. ECHA: *The use of alternatives to testing on animals for the REACH regulation*. 2017 [Online]. Available: https://echa.europa.eu/documents/10162/13639/alternatives_test_animals_2017_en.pdf.
5. USEPA: *Letter to stakeholders on EPA office of pesticide program's goal to reduce animal testing*. 2016 [Online]. Available: https://www.regulations.gov/document?D=EPA-HQ-OPP-2016-0093-0003. Accessed 9 August 2018.
6. USEPA: *Final strategic plan to promote development and implementation fo alternative test methods supporting toxic substances control act. Final strategic plan to promote development and implementation fo alternative test methods supporting toxic substances cont*. 2018 [Online]. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2017-0559-0903. Accessed 9 August 2018.
7. USFDA: *FDA's predictive toxicology roadmap*. 9 August 2018 [Online]. Available: https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/RegulatoryScience/UCM587831.pdf [Accessed 2018].
8. Bishop PL, Willett CE: **The use and acceptance of other scientifically relevant information (OSRI) in the U.S. Environmental protection agency (EPA) endocrine disruptor screening program**. *Birth Defects Res Part B Dev Reproductive Toxicol* 2013, **101**:3–22.
9. ** Paparella M: **Uncertainties of testing methods: what do we (want to) know about carcinogenicity?** *ALTEX* 2017:235–252.
10. Hoffmann S, Kleinstreuer N, Alépée N, Allen D, Api AM, Ashikaga T, Clouet E, Cluzel M, Desprez B, Gellatly N, Goebel C, Kern PS, Klaric M, Kühnl J, Lalko JF, Martinozzi-Teissier S, Mewes K, Miyazawa M, Parakhia R, van Vliet E, Zang Q, Petersohn D: **Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database**. *Crit Rev Toxicol* 2018, **48**:344–358.
11. ** Cohen SM: **The relevance of experimental carcinogenicity studies to human safety**. *Curr Opin Toxicol* 2017, **3**:6–11.
12. Stewart J: **The relevance of experimental reproductive studies in safety assessment**. *Curr Opin Toxicol* 2017, **3**:30–39.
13. Cohen SM, Arnold LL: **Chemical carcinogenesis**. *Toxicol Sci* 2010, **120**(Supplement 1):S76–S92.
14. Hoffmann S: *LLNA variability: an essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies*. ALTEX; 2015.
15. Sonawane KB, Cheng N, Hansen RA: **Serious adverse drug events reported to the FDA: analysis of the FDA adverse event reporting system 2006-2014 database**. *J Manag Care Spec Pharm* 2018, **24**:682–690.
16. Kleinstreuer NC, Ceger PC, Allen DG, Strickland J, Chang X, Hamm JT, Casey WM: **A curated database of rodent uterotrophic bioactivity**. *Environ Health Perspect* 2016, **124**:556–562.
17. ** Luechtefeld T, Marsh D, Rowlands C, Hartung T: **Machine learning of toxicological big data enables read-across structure activity relationships (RASAR) outperforming animal test reproducibility**. *Toxicol Sci* 2018, **165**:198–212.

18. Smith MT, Guyton KZ, Gibbons CF, Fritz JM, Portier CJ, Rusyn I, DeMarini DM, Caldwell JC, Kavlock RJ, Lambert PF, Hecht SS, Bucher JR, Stewart BW, Baan RA, Cogliano VJ, Straif K: **Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis**. *Environ Health Perspect* 2016, **124**:713–721.

19. * Guyton KZ, Rieswijk L, Wang A, Chiu WA, Smith MT: **Key characteristics approach to carcinogenic hazard identification**. *Chem Res Toxicol* 2018, **31**:1290–1292.

20. Chiu W: *Use of high-throughput in vitro toxicity screening data in cancer hazard evaluations by IARC Monograph Working Groups*. ALTEX; 2018:51–64.

21. * Guyton KZ, Rusyn I, Chiu WA, Corpet DE, van den Berg M, Ross MK, Christiani DC, Beland FA, Smith MT: **Application of the key characteristics of carcinogens in cancer hazard identification**. *Carcinogenesis* 2018, **39**:614–622.

22. OECD: **Test No. 442D: in vitro skin sensitisation: ARE-Nrf2 luciferase test method**. In *OECD guidelines for the testing of chemicals, section 4*. Paris: OECD Publishing; 2018.

23. OECD: **Test No. 442E: *In vitro* skin sensitisation: *In vitro* skin sensitisation assays addressing the key event on activation of dendritic cells on the adverse outcome pathway for skin sensitisation**. In *OECD guidelines for the testing of chemicals, section 4*. Paris: OECD Publishing; 2018.

24. OECD: **Test No. 439: in vitro skin irritation: reconstructed human epidermis test method**. In *OECD guidelines for the testing of chemicals, section 4*. Paris: OECD Publishing; 2015.

25. OECD: **Test No. 492: reconstructed human Cornea-like Epithelium (RhCE) test method for identifying chemicals not requiring classification and labelling for eye irritation or serious eye damage**. In *OECD guidelines for the testing of chemicals, section 4*. Paris: OECD Publishing; 2018.

26. ECHA: *Guidance on the biocidal products regulation. Volume III: human health. Part A: information requirements*. 2018 [Online]. Available: https://echa.europa.eu/documents/10162/23036412/bpr_guidance_vol_iii_part_a_en.pdf/05e4944d-106e-9305-21ba-f9a3a9845f93.

27. UNGHS: *Report to the sub-committee of experts on teh globally harmonized system of classification and labelling of chemicals on its thirty-fifth session*. 2018. https://www.unece.org/fileadmin/DAM/trans/doc/2018/dgac10c4/ST-SG-AC10-C4-70e.pdf.

28. Spanogiannopoulos P, Bess EN, Carmody RN, Turnbaugh PJ: **The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism HHS Public Access**. *Nat Rev Microbiol* 2016, **14**:273–287.

29. Baell JB, Holloway GA: **New substructure filters for removal of Pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays**. *J Med Chem* 2010, **53**:2719–2740.

30. Bruns RF, Watson IA: *Rules for identifying potentially reactive or promiscuous compounds*. 2012.

31. OECD: **Guidance document on good *In vitro* method practices (GIVIMP)**. In *OECD Series on Testing and assessment*, vol. 286. Paris: OECD Publishing; 2018.

32. Luechtefeld T: *Global analysis of publicly available safety data for 9,801 substances registered under REACH from 2008-2014*. ALTEX; 2016.

33. Aschner M: *Reference compounds for alternative test methods to indicate developmental neurotoxicity (DNT) potential of chemicals: example lists and criteria for their selection and use*. ALTEX; 2016.

34. Judson R, Houck K, Martin M, Richard AM, Knudsen TB, Shah I, Little S, Wambaugh J, Setzer RW, Kothya P, Phuong J, Filer D, Smith D, Reif D, Rotroff D, Kleinstreuer N, Sipes N, Xia M, Huang R, Crofton K, Thomas RS: **Analysis of the effects of cell stress and cytotoxicity on *In vitro* assay activity across a diverse chemical and assay space**. *Toxicol Sci* 2016, **152**:323–339.

35. Judson RS, Magpantay FM, Chickarmane V, Haskell C, Tania N, Taylor J, Xia M, Huang R, Rotroff DM, Filer DL, Houck KA, Martin MT, Sipes N, Richard AM, Mansouri K, Setzer RW, Knudsen TB, Crofton KM, Thomas RS: **Integrated model of chemical perturbations of a biological pathway using 18 *In vitro* high-throughput screening assays for the estrogen receptor**. *Toxicol Sci Offic J Soc Toxicol* 2015, **148**:137–154.

36. Dahlin JL, Baell J, Walters MA: *Assay interference by chemical reactivity*. Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004.

37. * OECD: **Detailed review paper on the use of metabolising systems for *In vitro* testing of endocrine disruptors**. In *OECD Series on Testing and assessment*, vol. 97. Paris: OECD Publishing; 2014.

38. Lilienblum W, Dekant W, Foth H, Gebel T, Hengstler JG, Kahl R, Kramer P-J, Schweinfurth H, Wollin K-M: **Alternative methods to safety studies in experimental animals: role in the risk assessment of chemicals under the new European Chemicals Legislation (REACH)**. *Arch Toxicol* 2008, **82**:211–236.

39. Jackson JP, Li L, Chamberlain ED, Wang H, Ferguson SS: **Contextualizing hepatocyte functionality of cryopreserved HepaRG cell cultures**. *Drug Metab Dispos* 2016, **44**:1463–1479.

40. DeGroot DE, Swank A, Thomas RS, Strynar M, Lee M-Y, Carmichael PL, Simmons SO: **mRNA transfection retrofits cell-based assays with xenobiotic metabolism**. *J Pharmacol Toxicol Methods* 2018, **92**:77–94.

41. Garcia-Canton C, Minet E, Anadon A, Meredith C: **Metabolic characterization of cell systems used in *in vitro* toxicology testing: lung cell system BEAS-2B as a working example**. *Toxicol In Vitro* 2013, **27**:1719–1727.

42. Pinto CL, Mansouri K, Judson R, Browne P: **Prediction of estrogenic bioactivity of environmental chemical metabolites**. *Chem Res Toxicol* 2016, **29**:1410–1427.

43. Heleno SA, Martins A, Queiroz MJR, Ferreira IC: **Bioactivity of phenolic acids: metabolites versus parent compounds: a review**. *Food Chem* 2015, **173**:501–513.

44. Ashraf MN, Asghar MW, Rong Y, Doschak MR, Kiang TKL: **Advanced *In vitro* HepaRG culture systems for xenobiotic metabolism and toxicity characterization**. *Eur J Drug Metab Pharmacokinet* 2018.

45. Ramaiahgari SC, Waidyanatha S, Dixon D, Devito MJ, Paules RS, Ferguson SS: **Three-dimensional (3D) HepaRG spheroid model with physiologically relevant xenobiotic metabolism competence and hepatocyte functionality for liver toxicity screening**. *Toxicol Sci* 2017, **159**:124–136.

46. Gouliarmou V, Lostia AM, Coecke S, Bernasconi C, Bessems J, Dorne JL, Ferguson S, Testai E, Remy UG, Brian Houston J, Monshouwer M, Nong A, Pelkonen O, Morath S, Wetmore BA, Worth A, Zanelli U, Zorzoli MC, Whelan M: **Establishing a systematic framework to characterise *in vitro* methods for human hepatic metabolic clearance**. *Toxicol Vitro* 2018, **53**:233–244.

47. Bessems JG, Loizou G, Krishnan K, Clewell Iii HJ, Bernasconi C, Bois F, Coecke S, Collnot E-M, Diembeck W, Farcal LR, Geraets L, Gundert-Remy U, Kramer N, Küsters G, Leite SB, Pelkonen OR, Schröder K, Testai E, Wilk-Zasadna I, Zaldívar-Comenges J-M: **PBTK modelling platforms and parameter estimation tools to enable animal-free risk assessment: recommendations from a joint EPAA â€“ EURL ECVAM ADME workshop**. *Regul Toxicol Pharmacol* 2014, **68**:119–139.

48. Bell SM, Chang X, Wambaugh JF, Allen DG, Bartels M, Brouwer KL, Casey WM, Choksi N, Ferguson SS, Fraczkiewicz G, Jarabek AM, Ke A, Lumen A, Lynn SG, Paini A, Price PS, Ring C, Simon TW, Sipes NS, Sprankle CS, Strickland J, Troutman J, Wetmore BA, Kleinstreuer NC: ***In vitro* to *in vivo* extrapolation for high throughput prioritization and decision making**. *Toxicol Vitro* 2018, **47**:213–227.

49. Wang D: **Infer the *in vivo* point of departure with ToxCast *in vitro* assay data using a robust learning approach**. *Arch Toxicol* 2018, **92**:2913–2922.

50. Abdullah R, Alhusainy W, Woutersen J, Rietjens IM, Punt A: **Predicting points of departure for risk assessment based on**

***in vitro* cytotoxicity data and physiologically based kinetic (PBK) modeling: the case of kidney toxicity induced by aristolochic acid I**. *Food Chem Toxicol* 2016, **92**:104–116.

51. Schmidt M, Raghavan B, Müller V, Vogl T, Fejer G, Tchaptchet S, Keck S, Kalis C, Nielsen PJ, Galanos C, Roth J, Skerra A, Martin SF, Freudenberg MA, Goebeler M: **Crucial role for human Toll-like receptor 4 in the development of contact allergy to nickel**. *Nat Immunol* 2010, **11**:814–819.

52. **Species differences in thyroid, kidney and urinary bladder carcinogenesis**. In *Proceedings of a consensus conference. Lyon, France, 3–7 November 1997*, vol. 147. IARC scientific publications; 1999:1–225.

53. International Agency for Research on Cancer: *Predictive value of rodent forestomach and gastric neuroendocrine tumours in evaluating carcinogenic risks to humans : views and expert opinions of an IARC working group, Lyon, 29 November-1 December 1999*. International Agency for Research on Cancer, World Health Organization; 2003:183.

54. Roberts DW, Api AM: **Chemical applicability domain of the local lymph node assay (LLNA) for skin sensitisation potency. Part 4. Quantitative correlation of LLNA potency with human potency**. *Regul Toxicol Pharmacol* 2018, **96**:76–84.

55. Amcoff P: *Unclassified ENV/JM/MONO(2005)14 joint meeting of the chemicals committee and the working party on chemicals, pesticides and biotechnology OECD series on testing and assessment number 34 guidance document on the validation and international acceptance of new or updated test methods for hazard assessment*. 2005.

56. Kleinstreuer NC, Hoffmann S, Alépée N, Allen D, Ashikaga T, Casey W, Clouet E, Cluzel M, Desprez B, Gellatly N, Göbel C, Kern PS, Klaric M, Kühnl J, Martinozzi-Teissier S, Mewes K, Miyazawa M, Strickland J, van Vliet E, Zang Q, Petersohn D: **Non-animal methods to predict skin sensitization (II): an assessment of defined approaches**. *Crit Rev Toxicol* 2018, **48**:359–374.

57. DeSesso JM: **Future of developmental toxicity testing**. *Curr Opin Toxicol* 2017, **3**:1–5.

58. ** Bossa C, Benigni R, Tcheremenskaia O, Battistelli CL: **(Q)SAR methods for predicting genotoxicity and carcinogenicity: scientific rationale and regulatory frameworks**. *Methods Mol Biol* 2018, **1800**:447–473.

59. 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*. 2010.

60. Leist M, Ghallab A, Graepel R, Marchan R, Hassan R, Hougaard Bennekou S, Limonciel A, Vinken M, Schildknecht S, Waldmann T, Danen E, van Ravenzwaay B, Kamp H, Gardner I, Godoy P, Bois FY, Braeuning A, Reif R, Oesch F, Drasdo D, Höhme S, Schwarz M, Hartung T, Braunbeck T, Beltman J, Vrieling H, Sanz F, Forsby A, Gadaleta D, Fisher C, Kelm J, Fluri D, Ecker G, Zdrazil B, Terron A, Jennings P, van der Burg B, Dooley S, Meijer AH, Willighagen E, Martens M, Evelo C, Mombelli E, Taboureau O, Mantovani A, Hardy B, Koch B, Escher S, van Thriel C, Cadenas C, Kroese D, van de Water B, Hengstler JG: **Adverse outcome pathways: opportunities, limitations and open questions**. *Arch Toxicol* 2017, **91**:3477–3505.

61. OECD: *Guidance document for the use of advers outcome pathways in developing integrated approaches to testing and assessment (IATA)*. OECD Series on Test Assessment No. 260 JT03407308. 2016.

62. Knudsen LE, Smith A, Törnqvist E, Forsby A, Tähti H: **Nordic symposium on "toxicology and pharmacology without animal experiments - will it be possible in the next 10 years? "**. *Basic Clin Pharmacol Toxicol* 2018.

63. 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 No. 256. 2016.

64. U. EPA. [Online]. Available: https://www.epa.gov/pesticides/epa-releases-draft-policy-reduce-animal-testing-skin-sensitization.

65. Ceridono M, Tellner P, Bauer D, Barroso J, Alépée N, Corvi R, De Smedt A, Fellows MD, Gibbs NK, Heisler E, Jacobs A, Jirova D, Jones D, Kandárová H, Kasper P, Akunda JK, Krul C, Learn D, Liebsch M, Lynch AM, Muster W, Nakamura K, Nash JF, Pfannenbecker U, Phillips G, Robles C, Rogiers V, Van De Water F, Liminga UW, Vohr HW, Wattrelos O, Woods J, Zuang V, Kreysa J, Wilcox P: **The 3T3 neutral red uptake phototoxicity test: practical experience and implications for phototoxicity testing - the report of an ECVAM-EFPIA workshop**. *Regul Toxicol Pharmacol* 2012.