

Letter

Potential of Concentration-Response Data to Broaden Regulatory Application of *In Vitro* Test Guidelines

Miriam N. Jacobs^{1, #}, Janine Ezendam^{2, #}, Betty Hakkert^{2, #} and Michael Oelgeschlaeger^{3, #}

¹Centre for Radiation, Chemical and Environmental Hazards, Public Health England, Chilton, OXON, UK; ²National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands; ³German Federal Institute for Risk Assessment, Berlin, Germany

The process for assessing the safety of chemicals comprises many steps and includes as a first step the hazard assessment of the chemical, i.e., the assessment of its intrinsic toxicological properties. For risk assessment purposes, this chemical-specific hazard data is then combined with chemical exposure estimations for human or environmental organisms to produce a chemical risk assessment with the goal of having a full understanding of the nature, magnitude, and probability of a potential adverse health or environmental effect of the chemical.

The international governmental platform for the development of hazard assessment tools for industrial chemicals is the Test Guideline Programme (TGP) of the Organisation for Economic Cooperation and Development (OECD). This generates consensually agreed test guidelines (TGs) that under the Mutual Acceptance of Data (MAD)¹ agreement provide a common basis for working in a harmonized manner, thereby reducing conflicting or duplicative requirements for regulation, saving economic costs, and greatly reducing experimental animal use in OECD countries. TG hazard assessment tools can be utilized for priority setting, risk assessment, and other activities within national or regional programmes.

In the early years of the OECD TGP, the programme was dominated by the development of animal (*in vivo*) TGs, as it was considered most relevant to conclude upon toxicological data from an intact animal model. *In vitro* test methodology was not so well progressed but was considered of value to prioritize chemicals to be tested subsequently *in vivo*.

Animal welfare concerns, particularly the 3Rs of replacing, reducing, and refining animal testing (Russell and Burch, 1959), whilst also advancing the quality and relevance of experimental techniques for humans, together with a global drive to advance

toxicity testing in the 21st century (NRC, 2007), have become fundamental to TG development, and in recent years a paradigm shift towards new approach methodology (NAM) has become evident in the international scientific community that is very much reflected in the OECD TGP. *In vitro* test method development is expanding enormously, and ways in which the generated data can be combined are being developed for specific toxicological endpoints and becoming the way forward for the regulatory safety testing of chemicals. The paradigm shift has led to the replacement of *in vivo* test methods by *in vitro* methods in integrated and tiered approaches in the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS) for the classification of skin corrosion and irritation (adopted December 2018), and for serious eye damage and eye irritation (adopted July 2021). UN GHS work is also ongoing to similarly update the chapter on skin sensitization testing, the latter following the publication of the OECD Defined Approach for Skin Sensitisation (DASS; Kleinstreuer et al., 2018; OECD, 2016a, 2021a).

Such approaches, now termed integrated approaches to testing and assessment (IATA), are structured approaches used for hazard identification (potential), hazard characterization (potency), and/or safety assessment (potential/potency and exposure) of a chemical or group of chemicals, which strategically integrate and weight all relevant data (weight of evidence: WoE) to inform regulatory decisions regarding potential hazard and/or risk and/or the need for further targeted testing. An IATA therefore optimizes and potentially reduces the number of tests that need to be conducted (OECD, 2016b). Ultimately, an IATA may be able to be developed into a defined approach (DA), i.e., an integrated testing approach consisting of a selection of information sources (e.g.,

[#] The authors are all national experts to the OECD Test Guideline Programme, and MNJ, BH, and MO are Human Health National Coordinators (for UK, NL, and Germany, respectively).

Received July 9, 2021; © The Authors, 2022.

ALTEX 39(2), 315–321. doi:10.14573/altex.2107091

Miriam N. Jacobs, PhD
Centre for Radiation, Chemical and Environmental Hazards,
Public Health England, Chilton, OXON OX11 0RQ, UK
(Miriam.Jacobs@phe.gov.uk)

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.

¹ <https://www.oecd.org/chemicalsafety/testing/mutualacceptanceofdatamad.htm> (accessed 07.11.2021)



in silico predictions, *in chemico*, and/or *in vitro* TG data) that are used in a specific combination such that the resulting data are interpreted using a fixed data interpretation procedure (DIP) (e.g., a mathematical or rule-based model) (OECD, 2016c). Such DAs are ultimately intended to completely replace animal testing for the respective endpoint.

Once an *in vitro/in chemico* TG is agreed and adopted at OECD, it is often assumed by policy makers that the TG is fit to be used in a variety of regulatory applications. However, many current *in vitro* TGs were originally developed and validated only for the purpose of screening/hazard identification. These *in vitro* test methods use prediction models to translate the *in vitro* data into a hazard classification of the test substance. The prediction models give dichotomous positive/negative answers by applying a pre-defined threshold, as determined and reported during the interlaboratory validation exercise. The threshold is based on experimental data separating reference chemicals known to be positive or negative for the measured activity based on animal or human data. However, the number of reference chemicals considered to be sufficient to determine such a threshold is not defined. Also, the process by which the threshold was derived is not currently included in the description of prediction models of *in vitro* and *in chemico* TGs. An inherent weakness of such a threshold is that distinguishing positive and negative chemicals with respect to the measured activity becomes more uncertain the closer the measured values approach this threshold. The definition of borderline ranges of uncertainty in which no clear conclusion can be drawn, as seen in the DASS (OECD, 2021a) and discussed in the guideline supporting information and by Kolle et al. (2021), may help to address this problem.

While multiple concentrations of reference chemicals are usually tested in such *in vitro* or *in chemico* assays by the developers, these data are often not used or interpreted in the prediction models of the *in vitro* assays. However, they provide a wealth of concentration-response information that may allow (relative) potency assessment of test compounds and/or the derivation of toxicological point(s) of departure (PoD) for risk assessment. A PoD is defined as the point on a toxicological dose-response curve established from experimental data or observational data corresponding to an estimated low or no effect level, most commonly the no-observed-adverse-effect level (NOAEL), lowest-observed-adverse-effect level (LOAEL) or statistical benchmark dose (BMD). PoD is closely related to the tipping point at which the chemical concentration being tested can trigger a shift from a molecular initiating event (MIE) or early key event (KE) of an adverse outcome pathway (AOP) into the subsequent KE, eventually leading to an adverse outcome. Thus, the PoD is the starting point from which one can extrapolate to, for instance, a toxicological reference dose (RfD), acceptable daily intake (ADI), or reference concentration (RfC). Our intention here is to raise the issue that neither concentration-response data nor descriptions of the process by which the threshold was derived are currently included in prediction models of *in vitro* and *in chemico* TGs. There is therefore a need to support and help to frame how overarching consistency in the data reporting of *in vitro* test method approaches being developed at the OECD can be achieved.

The dichotomous classification of substances is an important and pragmatic approach to prioritize a vast number of substances. It is essential to have such classification to enable clear-cut decisions for the consumer, transport, worker safety, etc., and also for subsequent follow-up. However, this also means that TGs only using such approaches cannot be used and were not designed to be used in scenarios where endpoint-specific concentration-response data are needed to provide relative potency information. Potency information is essential input for the use of assay data in approaches that are geared towards quantitative hazard and risk assessment, such as IATAs, DAs, quantitative AOPs, quantitative structure activity relationships (QSAR), and *in silico* modelling. This was recently recognized by the EURL ECVAM Scientific Committee (ESAC), which recommended possibilities should be sought for maximizing the use of concentration-response data in the assessment of the relevance of *in vitro* assays (ESAC, 2020).

In addition, to address this issue, we also need to clarify and define the role of weight of evidence (WoE) in point of departure prediction (PoD). In developing an IATA, WoE involves weighting the relative contributions of the individual *in vivo*, *in vitro* and *in silico* data to achieve a transparent understanding of the mechanistic (or mode of action) contributions to the overall resultant chemical hazard assessment as part of an integrated approach. Here, the assays and information sources are not fixed, as compared to a DA. As such, an IATA is more flexible than a DA. However, once an IATA is included into regulatory guidance documents, it may also be expected to follow a dichotomous decision tree (if-then) in a similar fashion to the process used for the UN GHS. Indeed, such clarification of the role of WoE in PoD prediction would help to facilitate harmonization between TGs and DAs for the UN GHS process.

Consequently, we propose exploring whether concentration-response data of existing *in vitro* OECD TGs can be used in prediction models and subsequently as inputs in the DIPs of DAs to cover both dichotomous answers and concentration-response, and to consistently expand the regulatory application of TGs beyond prioritization purposes. The intention is to initiate this discussion among regulators but also to encourage test method developers who are developing reliable and relevant methods for TG purposes to provide concentration-response data so that the data needed to develop respective future IATA and DA using their TG will be available.

Selected endpoint illustrations

Skin sensitization

The recently adopted DASS (OECD, 2021a) includes one DA, the 2o3 DA, to distinguish sensitizers from substances with no sensitizing activity. It also includes two separate DAs that predict skin sensitization CLP subcategories 1A and 1B, i.e., the ITSv1 and ITSv2. This subcategorization is based on concentration-response data derived from the individual assays (DPRA and h-CLAT) by applying respective predefined thresholds (Takanouchi et al., 2015). The procedure by which such threshold values are set is not defined, and during the evaluation of this DA, the prediction model needed to be changed because the scoring system used in the assays did not perform well for the strong (Cat

1A) skin sensitizers. This nicely exemplifies that it is problematic to define a threshold based on a particular set of data that might later prove to be unsuitable when the data set is expanded.

In the area of skin sensitization, the value of applying concentration-response information for skin sensitizer potency or PoD prediction is being explored in other DAs as well. Several of these DAs use *in vitro* test method data as inputs without first applying the TG's prediction model. For example, cysteine or lysine depletion data from the Direct Peptide Reactivity Assay (DPRA; OECD TG 442C; OECD, 2021b) are used as an input for the Skin Allergy Risk Assessment (SARA) DA (Reynolds et al., 2019; Gilmour et al., 2020), but instead of the OECD TG 442C cut-off value, a Bayesian probabilistic approach incorporating exposure parameters is utilized. The Bayesian approach reduces uncertainty by using random variables to model all sources of uncertainty in statistical models, including uncertainty resulting from lack of information. The Bayes formula also facilitates the sequential use of data as more data become available. A two-year collaboration between Unilever and NICEATM will further develop the SARA model for skin sensitization to expand the model database and functionality². The intention is to reduce uncertainties by improving the data quality, chemical applicability domains, and prediction models by developing more complex but also more accurate DAs for skin sensitization. This project is under consideration for inclusion in the OECD TGP workplan for 2022.

The development of the kinetic DPRA, which is now included in OECD TG 442C (OECD, 2021b), shows that more time points are needed compared to the original DPRA for greater accuracy in potency prediction. However, data from the original DPRA are currently used to subcategorize skin sensitizers in several DAs. This suggests that the inputs that are currently being used for potency prediction may not be optimal for this purpose and that both the kinetic DPRA and work in progress with respect to the SARA model, for example, could greatly improve this situation.

Cloud-based machine learning approaches are now also being integrated into draft TGs. The Genomic Allergen Rapid Detection (GARD™) assay uses a novel genomic biomarker signature that discriminates between weak and strong skin sensitizers (Gradin et al., 2020). The prediction model is constantly optimized as new data is generated, and the EURL ECVAM Scientific Advisory committee (ESAC) peer review³ was able to support progression of GARDskin towards a TG, but a need for further work for GARDpotency was identified.

Different mathematical model approaches and test data sets can be used to test the evolving models. Practical lessons learned from the DASS project were that a combination of methods can reduce uncertainty (if they confirm each other, results have non-overlapping limitations, and they address distinct KEs) (OECD, 2021a).

These insights highlight that additional information may be derived from *in vitro* OECD TGs for skin sensitizer potency or PoD prediction, thus broadening their use when concentration-response data is considered directly. Much data has been generat-

ed for the individual *in vitro* and *in silico* skin sensitization assays which will allow analysis of the value of the different parameters for the prediction of hazard and potency. Such analysis would be helpful in addressing the following issues:

- How can we better understand the dynamic response to determine the critical threshold point (the original lowest point above the noise) to derive the PoD?
- Could better use of concentration-response data aid in the definition of borderline ranges or reduce the uncertainty of dichotomous decisions (Kolle et al., 2019, 2021; Leontaridou et al., 2019; Gabbert et al., 2020)?
- Can we better understand which parameters/DIPs have the greatest impact/influence on DA accuracy by considering concentration-response information?
- Can we use these insights to identify new or optimize existing *in vitro* or *in chemico* assays (e.g., DPRA vs. kinetic DPRA) or to reduce *in vitro* TG variability?
- How can we address the challenges of interchangeability of *in vitro* tests around different key events within a DA where cut-offs used between *in vitro* assays are different, as for example with the U-SENS and h-CLAT?
- What is the most valuable quantitative parameter from a mechanistic view, not simply looking at the threshold? For example, for the U-SENS, CD86 is the principal marker, whilst for the h-CLAT, both CD86 and CD54 are utilized.

Phototoxicity

Phototoxicity is determined by comparing the relative potency of a substance in the absence or presence of UV light. One prediction model is based on the use of a photo-irritancy factor (PIF) that is defined as the ratio of the EC₅₀ values from UV-treated and untreated cells. An alternative prediction model, the mean photo effect (MPE), was developed by Peters and Holzhütter (2002) to detect phototoxic chemicals in the 3T3 NRU test (OECD TG 432; OECD 2019a,b). The MPE provides a more comprehensive analysis by comparing the area under the curve of UV-treated and untreated cells. While the MPE and PIF generally result in comparable predictivity values, the MPE enables a quantitative analysis of those concentration-responses that do not allow the derivation of a reliable EC₅₀ value if, for example, the highest concentration does not reduce viability sufficiently.

Endocrine assays

Currently, OECD TG 458, the *in vitro* human androgen receptor transcriptional activation assay for the detection of androgenic agonist and antagonist activity, and other methodologically similar endocrine-related OECD TGs rely on classifiers (e.g., percent concentration activation, such as PC₁₀ or PC₅₀, or effective concentration or inhibition concentration, such as EC₅₀, IC₅₀) to identify positive and negative chemicals irrespective of concentration-response data. A classifier is any algorithm that sorts data into labelled classes or categories of information. Although perti-

² <https://ntp.niehs.nih.gov/whatwestudy/niceatm/iccvm/agency-activities/index.html> (accessed 30.05.2021)

³ https://tsar.jrc.ec.europa.eu/system/files/Published/ESAC_Opinion_2021-01_GARD_%28online%29.pdf



Tab: 1: Summary of molecular and transformation phenotypic endpoints in a CTA BALB/c 3T3 transformics model following chemical treatment with 3-MCA at two concentrations

Mascolo et al. (2018). Permission to re-use for non-commercial reasons from the authors.

3-MCA	Main molecular endpoints			Transformation phenotypic endpoint
	24 h	72 h	32 d	32 d
0.04 µg/mL	– Cell adhesion mechanisms – Cell cycle regulation – Apoptosis regulation – Retinol metabolism	– Immune response – Apoptosis regulation – Estradiol metabolism	– No statistically significant gene modulation	None
4 µg/mL	– Immune response – Cytoskeleton remodeling – Cell cycle regulation – Apoptosis regulation – Retinol metabolism	– Immune response – Cell cycle regulation – Estradiol metabolism	– Immune response	Cell transformation with evident malignant foci, sustained by the immune response

ment for a dichotomous classification, this approach neglects other high-quality and highly reproducible information that may be attained from the full concentration-response curve. This is acknowledged in the respective ESAC peer review, which states that the use of the concentration-response data could be maximized in the assessment of assays (ESAC, 2020).

Both OECD TGs 455 and 458, the *in vitro* ER and AR transactivation assays (ERTA and ARTA respectively), particularly the recently validated AR-CALUX[®], provide superb concentration-response data on one of the broadest chemical applicability domains (with 46 representative test chemicals) that have so far been included in TGs developed within the OECD Validation Management Group-Non Animal (VMG-NA). However, for TG 458, this is not being built into the prediction model, as the older dichotomous TG model is followed to create one performance-based TG that includes three different assays (OECD, 2020; Milcamps et al., 2020; Park et al., 2021) as opposed to a TG that includes only one test method. This is a lost opportunity as the original concentration-response data could facilitate building improved *in silico* statistical models such as QSAR and developing IATA applications. With respect to the ARTAs in TG 458, a suggested approach is to include logPC₁₀ and logIC₃₀ values together with standard deviation (SD) values along with the positive/negative calls.

Non-genotoxic carcinogen IATA

Transcriptomic analyses of different chemical concentrations can assist greatly in understanding different mechanistic responses, not only for *in vivo* studies (Thomas et al., 2012) but also within more complex *in vitro* assays as shown, for example, with cell transformation assays (CTAs) to identify non-genotoxic carcinogens. Using a high-throughput microarray approach, key infor-

mation on molecular mechanisms related to the final adverse outcome represented by formation of malignant foci was reported by Mascolo et al. (2018) for the BALB/c 3T3 CTA. In order to improve the use of this assay in the integrated testing strategy for carcinogenesis, a method termed transformics, which combines the CTA and transcriptomics to identify the molecular steps leading to *in vitro* malignant transformation, was developed.

3-methylcholanthrene (3-MCA) is a recognized genotoxic chemical that is also able to induce *in vitro* cell transformation via non-genotoxic mechanisms. It was studied at both transforming and sub-transforming (i.e., a lower concentration than that at which full transformation is first observed, but at which a reversible response is seen) concentrations over different exposure periods in BALB/c 3T3 cells. At 0.04 µg/mL (low dose) and 4 µg/mL (high dose) 3-MCA, transient or persistent mechanistic differences, respectively, were evident at different time points (24 h, 72 h and 32 days), with no cytotoxicity, as shown in Table 1.

The table shows the main biological targets for each concentration and time for 3-MCA. The immune response is flagged in bold, as it was the response that persisted across time and was not reversed at the higher concentration of 4 µg/mL 3-MCA. At 24 h, both concentrations modulated cell cycle, apoptosis, and retinol metabolism regulation. Whilst the cell adhesion mechanism modulation was associated only with 0.04 µg/mL 3-MCA as a cell adaptive response, the cytoskeleton remodeling observed at 4 µg/mL 3-MCA was the early KE towards cell transformation. At 72 h, the immune response became the distinctive trait of the gene modulation for both concentrations, but it was associated with transcriptional modulation of the apoptotic processes at 0.04 µg/mL and with alteration of cell-cycle regulation at 4 µg/mL 3-MCA. This, therefore, led to a notably different fate for the two cell populations. At 32 days, no significant modulation was

observed at 0.04 $\mu\text{g/mL}$ 3-MCA, while the phenotypic outcome of the cell transformation (malignant foci), still sustained by the immune response, was visibly evident in the plates treated with 4 $\mu\text{g/mL}$ 3-MCA.

The results gave evidence for a potential key role of the immune system, and it was postulated that the aryl hydrocarbon receptor (AhR) pathway may also be part of the initial steps of the *in vitro* transformation process induced by 3-MCA (Mascolo et al., 2018), suggesting that, in the CTA, the initiating and promoting events are related to non-genotoxic mechanisms. Similar results have been obtained for a further two chemicals by the same laboratory (manuscripts in preparation), although the work has not yet been reproduced in an independent laboratory.

Here, the combination of omics tools with concentration-response information over different time points has elucidated likely mechanisms within the CTA that could allow the use of these assays in a non-genotoxic carcinogen IATA (Jacobs et al., 2020).

What do we want to see, and how do we get there?

We have presented four very different endpoint scenarios for which concentration-response information contained in the assay data is of high utility and for which there is already some recognition. Indeed, ESAC peer review of some of these *in vitro* test method validation efforts have noted the essentiality of these approaches for quantitative hazard and risk assessment such as IATA.

For each *in vitro* assay that could provide input for potency assessment, it is strongly recommended that test method developers provide a documented, transparent analysis of all experimental data to the OECD to facilitate an understanding of the dynamic response and potential borderline ranges that are required to determine the critical threshold from which a PoD may be derived for a given chemical (as seen with 3-MCA concentrations in Mascolo et al. (2018) for example). Concentration-response curves, as needed for the identification of the intrinsic hazard properties, also need to be fit for derivation of PoD for risk assessment. In this way, the OECD TGP and regulators can better understand how the quantitative parameters can be employed for PoD determination and assessment of potency.

We need to know which parameters are the most informative to derive a PoD and if the correct cut-offs are used in the prediction model/DIP, and we can do this most effectively if we have the concentration-response data from these assays to enable subsequent statistical analyses and bioinformatics design for IATAs. We would then be able to build this information into the relevant TG, and/or subsequent DA and/or IATA.

Existing guidance on developing *in vitro* methods is already available in, for example, the OECD Guidance Document on Good *In Vitro* Practices (GIVIMP) (OECD, 2018) and needs to be consulted during the process of test method development. TGs have OECD Harmonized Templates to support the reporting of (mainly *in vivo*) TG results, which use existing open data standard formats that could be adapted for concentration-response in-

formation from *in vitro* methods in addition to threshold and dichotomous DIPs. Further, assay templates have been developed to collect such information (Tab. S1⁴), and reporting guidance is in development for QSARs and *in silico* tools as well as for omics data (Harrill et al., 2021). How representative a parameter is for an AOP/IATA/DA from a mechanistic point of view is a relevant question, but out of the scope of this paper, and would need to be addressed subsequently in the relevant quantitative AOP and/or IATA endpoint development.

Up to this point, we have focused on standard test chemicals. Additional considerations are needed to address, for instance, chemical mixtures, polymers, advanced materials, unknown variable composition or biological substances (UVCBs), nanomaterials, and others. For example, the plethora of nanomaterials and the very limited information on potential reference nanomaterials is a major problem. Nanomaterials employed at low concentrations are likely to have a good dispersion, but at higher concentrations agglomeration may lead to false-negative results. A harmonized, agreed set of nanomaterials would be very helpful, and the forthcoming OECD preliminary guidance on addressing nanomaterials in *in vitro* genotoxicity testing will provide some initial support.

Key questions to address

In framing this appeal for the inclusion of concentration-response information in the future development of *in vitro* OECD TGs, there are several possible routes that can be taken to facilitate increased functionality and utility of the TGs in the development of IATAs and DAs.

This could start in the short term with the TGs/assays that are intended for adoption in 2022 or that have been adopted recently, such as the DASS (OECD, 2021a) and the ARTA (Park et al., 2021; OECD, 2020). Current OECD IATA projects, such as that for non-genotoxic carcinogens (Jacobs et al., 2020), can examine the best way to adapt relevant current *in vitro* TGs to be included in the IATA and can therefore provide relevant examples that can be translated to other complex IATA endpoints.

The OECD working group of National Coordinators to the TGP could consider that:

1. Test method developers should be requested to provide quantitative concentration-response curves in addition to regulatory requirements of thresholds for dichotomous interpretation as the draft TGs are developed and validated. Comparative tabulation of the different PoD and quantitative parameters should be performed for each TG/DA.
2. Where possible, an uncertainty analysis can inform on the most relevant parameters/inputs to be used. This will need to include transparent and sustainable storage of all experimental data used for threshold definitions according to FAIR principles (OECD, 2019b). In addition, an uncertainty analysis of *in vivo* reference data as, for example, conducted for the rodent cancer bioassay (Paparella et al., 2017), can be useful.

⁴ doi:10.14573/altex.2107091s



3. The highest concentration of test chemical used should be confirmed/checked by cytotoxicity testing. Where no cytotoxicity is detected, the concentration-response levels are well below any effect, and solubility issues are adequately addressed, the highest concentrations used for testing might not be adequate. Metabolites of test chemicals also need to be addressed by combination testing with validated *in vitro* metabolism assays such as the HepRG™ CYP induction test method (Bernasconi et al., 2019).

In the longer term, it would be appropriate to consider creating an easily searchable electronic validation data repository (i.e., in addition to the OECD Series in Testing and Assessment validation reports) at the OECD that might be similar in structure to, for example, Metapath (Kolanczyk et al., 2012).

Our general aim discussed herein is to point out that for evolving regulatory needs we need to have comprehensive *in vitro* test method data available in a format that allows the combination of data from distinct methods for the subsequent development of IATAs or DAs. Here we have only discussed standard chemicals but not the considerations needed specifically to address different forms, such as nanomaterials, which have their own challenges. We hope that this discussion will encourage test method developers and regulators to consider the points raised here to facilitate improved regulatory take-up, acceptance, and use of NAMs.

References

- Bernasconi, C., Pelkonen, O., Andersson, T. B. et al. (2019). Validation of *in vitro* methods for human cytochrome P450 enzyme induction: Outcome of a multi-laboratory study. *Toxicol In Vitro* 60, 212-228. doi:10.1016/j.tiv.2019.05.019
- ESAC (2020). Opinion on the Scientific Validity of the AR-CALUX® Test Method. EUR 30272 EN, Publications Office of the European Union, Luxembourg. doi:10.2760/885798
- Gabbert, S., Mathea, M., Kolle, S. N. et al. (2020). Accounting for precision uncertainty of toxicity testing: methods to define borderline ranges and implications for hazard assessment of chemicals. *Risk Anal*, online ahead of print. doi:10.1111/risa.13648
- Gilmour, N., Kern, P., Alépée, N. et al. (2020). Development of a next generation risk assessment framework for the evaluation of skin sensitisation of cosmetic ingredients. *Regul Toxicol Pharmacol* 116, 104721. doi:10.1016/j.yrtph.2020.104721
- Gradin, R., Johansson, A., Forreryd, A. et al. (2020). The GARD potency assay for potency-associated subclassification of chemical skin sensitizers-rationale, method development, and ring trial results of predictive performance and reproducibility. *Toxicol Sci* 176, 423-432. doi:10.1093/toxsci/kfaa068
- Harrill, J. A., Viant, M. R., Yauk, C. L. et al. (2021). Progress towards an OECD reporting framework for transcriptomics and metabolomics in regulatory toxicology. *Regul Toxicol Pharmacol* 125, 105020. doi:10.1016/j.yrtph.2021.105020
- Jacobs, M. N., Colacci, A., Corvi, R. et al. (2020). Chemical carcinogen safety testing: OECD expert group international consensus on the development of an integrated approach for the testing and assessment of chemical non-genotoxic carcinogens. *Arch Toxicol* 94, 2899-2923. doi:10.1007/s00204-020-02784-5
- Kleinstreuer, N. C., Hoffmann, S., Alépée, N. et al. (2018). Non-animal methods to predict skin sensitization (II): An assessment of defined approaches. *Crit Rev Toxicol* 48, 359-374. doi:10.1080/10408444.2018.1429386
- Kolanczyk, R. C., Schmieder, P., Jones, W. J. et al. (2012). Meta-Path: An electronic knowledge base for collating, exchanging and analysing case studies of xenobiotic metabolism. *Regul Toxicol Pharmacol* 63, 84-96. doi:10.1016/j.yrtph.2012.02.013
- Kolle, S. N., Natsch, A., Gerberick, G. F. et al. (2019). A review of substances found positive in 1 of 3 *in vitro* tests for skin sensitization. *Regul Toxicol Pharmacol* 106, 352-368. doi:10.1016/j.yrtph.2019.05.016
- Kolle, S. N., Mathea, M., Natsch, A. et al. (2021). Assessing experimental uncertainty in defined approaches: Borderline ranges for *in chemico* and *in vitro* skin sensitization methods determined from ring trial data. *Appl In Vitro Toxicol* 7, 102-111. doi:10.1089/aivt.2021.0003
- Leontaridou, M., Gabbert, S. and Landsiedel, R. (2019). The impact of precision uncertainty on predictive accuracy metrics of non-animal testing methods. *ALTEx* 36, 435-446. doi:10.14573/altex.1810111
- Mascolo, M. G., Perdichizzi, S., Vaccari, M. et al. (2018). The transformics assay: First steps for the development of an integrated approach to investigate the malignant cell transformation *in vitro*. *Carcinogenesis* 39, 955-967. doi:10.1093/carcin/bgy037
- Milcamps, A., Liska, R., Langezaal, L. et al. (2020). Validation Study Report: Performance assessment of the AR-CALUX® *in vitro* method. EUR 30044 EN, Publications Office of the European Union, Luxembourg, 2020. doi:10.2760/89293
- NRC – National Research Council (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC, USA: The National Academies Press. doi:10.17226/11970
- OECD (2016a). Guidance Document 256. 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; Annex 1; Annex 2 ENV/JM/MONO(2016)29 3. *OECD Environment, Health and Safety Publications Series on Testing and Assessment, No. 256*.
- OECD (2016b). Guidance Document for the use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment. *Series on Testing and Assessment, No. 260*. OECD, Paris. [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2016\)67&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2016)67&doclanguage=en)
- OECD (2016c). Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment. *Series on Testing and Assessment, No. 255*. OECD, Paris. <http://www.oecd.org/officialdocuments/>

- publicdisplaydocumentpdf/?cote=env/jm/mono(2016)67&doclanguage=en
- OECD (2018). Guidance Document on Good In Vitro Method Practices (GIVIMP). *OECD Series on Testing and Assessment, No. 286*. OECD Publishing, Paris. doi:10.1787/9789264304796-en
- OECD (2019a). Test No. 432: In Vitro 3T3 NRU Phototoxicity Test. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264071162-en
- OECD (2019b). Guiding Principles on Good Practices for the Availability/Distribution of Protected Elements in OECD Test Guidelines. *Series on Testing and Assessment, No. 298*. OECD, Paris. [https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV-JM-MONO\(2019\)14%20&doclanguage=en](https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV-JM-MONO(2019)14%20&doclanguage=en)
- OECD (2020). Test No. 458: Stably Transfected Human Androgen Receptor Transcriptional Activation Assay for Detection of Androgenic Agonist and Antagonist Activity of Chemicals. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264264366-en
- OECD (2021a). Guideline No. 497: Defined Approaches on Skin Sensitisation. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/b92879a4-en
- OECD (2021b). Test No. 442C: In Chemico Skin Sensitisation: Assays addressing the Adverse Outcome Pathway Key Event on Covalent Binding to Proteins. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264229709-en
- Paparella, M., Colacci, A. and Jacobs, M. N. (2017). Uncertainties of testing methods: What do we (want to) know about carcinogenicity? *ALTEX* 34, 235-252. doi:10.14573/altex.1608281
- Park, Y., Jung, D. W., Milcamps, A. et al. (2021). Characterisation and validation of an in vitro transactivation assay based on the 22Rv1/MMTV_GR-KO cell line to detect human androgen receptor agonists and antagonists. *Food Chem Toxicol* 152, 112206. doi:10.1016/j.fct.2021.112206
- Peters, B. and Holzhütter, H. G. (2002). In vitro phototoxicity testing: Development and validation of a new concentration response analysis software and biostatistical analyses related to the use of various prediction models. *Altern Lab Anim* 30, 415-432. doi:10.1177/026119290203000405
- Reynolds, J., MacKay, C., Gilmour, N. et al. (2019). Probabilistic prediction of human skin sensitiser potency for use in next generation risk assessment. *Comput Toxicol* 9, 36-49. doi:10.1016/j.comtox.2018.10.004
- Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. Free digital edition, Johns Hopkins Center for Alternatives to Animal Testing (CAAT). <https://caat.jhsph.edu/principles/the-principles-of-humane-experimental-technique>
- Takenouchi, O., Fukui, S., Okamoto, K. et al. (2015). Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. *J Appl Toxicol* 35, 1318-1332. doi:10.1002/jat.3127
- Thomas, R. S., Clewell, H. J. 3rd, Allen, B. C. et al. (2012). Integrating pathway-based transcriptomic data into quantitative chemical risk assessment: A five chemical case study. *Mutat Res* 746, 135-1343. doi:10.1016/j.mrgentox.2012.01.007

Conflict of interest

The authors have no conflicts of interests.