



Concept Article

Animal Metrics: Tracking Contributions of New Approach Methods to Reduced Animal Use

M. Sue Marty¹, Amanda K. Andrus¹ and Katherine A. Groff²

¹Dow, Inc., Toxicology and Environmental Research and Consulting, Midland, MI, USA; ²People for the Ethical Treatment of Animals, Norfolk, VA, USA

Abstract

Many companies and global regulatory programs have expressed the intent to move away from *in vivo* animal testing to new approach methods (NAMs) as part of product safety assessments. NAMs, which include non-animal approaches for testing and assessment – from computer-based modeling to *in chemico* or *in vitro* models – allow faster data generation with potentially greater relevance to humans while avoiding animal use. To monitor progress implementing NAMs, each organization first must define what is in scope, starting with the definition of “animal” (e.g., mammals, vertebrates) and applicable studies (e.g., animals used for “in-house” experiments, at contract research organizations, as part of environmental monitoring). Next, organizations must establish baseline animal use, including defined rules for inclusion/exclusion of animals that ensure consistency in future assessments. Lastly, organizations must establish metrics for animal savings based on the utility of NAM data. This paper presents one approach to establish “animal use” metrics in a toxicology program at The Dow Chemical Company. The premise of our program is that most NAM information has value for animal savings, but the value depends on how data are used (e.g., research and development, screening, or regulatory requirements) and the level of certainty for internal decision-making. This manuscript provides metrics on the impact of NAMs, allowing a quantitative assessment of animal use numbers over time, accountability for resources spent on NAM development, and identification of areas where NAM development is still needed. This approach can be refined for use at other organizations.

1 Introduction

Historically, toxicology has relied on animal-based studies to characterize potential toxicity hazards and risks to humans. These data have been used by regulatory agencies to determine whether a chemical can be used safely; however, use of animals for safety assessments is expensive, time-consuming, raises ethical issues, and is increasingly scrutinized for relevance to human health outcomes. In 2007, the National Research Council of the National Academies (NRC, 2007) published a report entitled “Toxicity Testing in the 21st Century: A Vision and a Strategy”, which fostered an evolution in toxicology away from animal-based testing to a pathway-based approach where non-animal models can be used to understand initial interactions of chemicals with target sites at the molecular, cellular and/or tissue level. Whenever possible, these non-animal approaches are based on the use of hu-

man cells/tissues to improve relevance to human health effects and limit the use of animals in testing.

Subsequently, several global and national regulatory and standards programs (e.g., US Environmental Protection Agency (EPA), US Food and Drug Administration (FDA), US Consumer Product Safety Commission (CPSC)) have announced plans to move away from animal testing to the extent possible either through the acceptance of new approach methods (NAMs; US EPA, 2018) – namely *in silico* and *in vitro* approaches – or through outright bans of animal-based tests (e.g., European Union Cosmetics Regulation). The US EPA had proposed a goal to eliminate animal testing in mammals by 2035, although this date has been stricken from the 2021 New Approach Methods Work Plan (US EPA, 2019a,b, 2021). In recent years, the Organisation for Economic Cooperation and Development (OECD) has developed internationally accepted test guidelines for several NAMs

Received July 21, 2021; Accepted October 19, 2021; Epub October 22, 2021; © The Authors, 2022.

ALTEX 39(1), 095-112. doi:10.14573/altex.2107211

Correspondence: M. Sue Marty
The Dow Chemical Company, Toxicology and Environmental Research
and Consulting
1803 Building, Midland, MI 48667, USA
(msmarty@dow.com)

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.

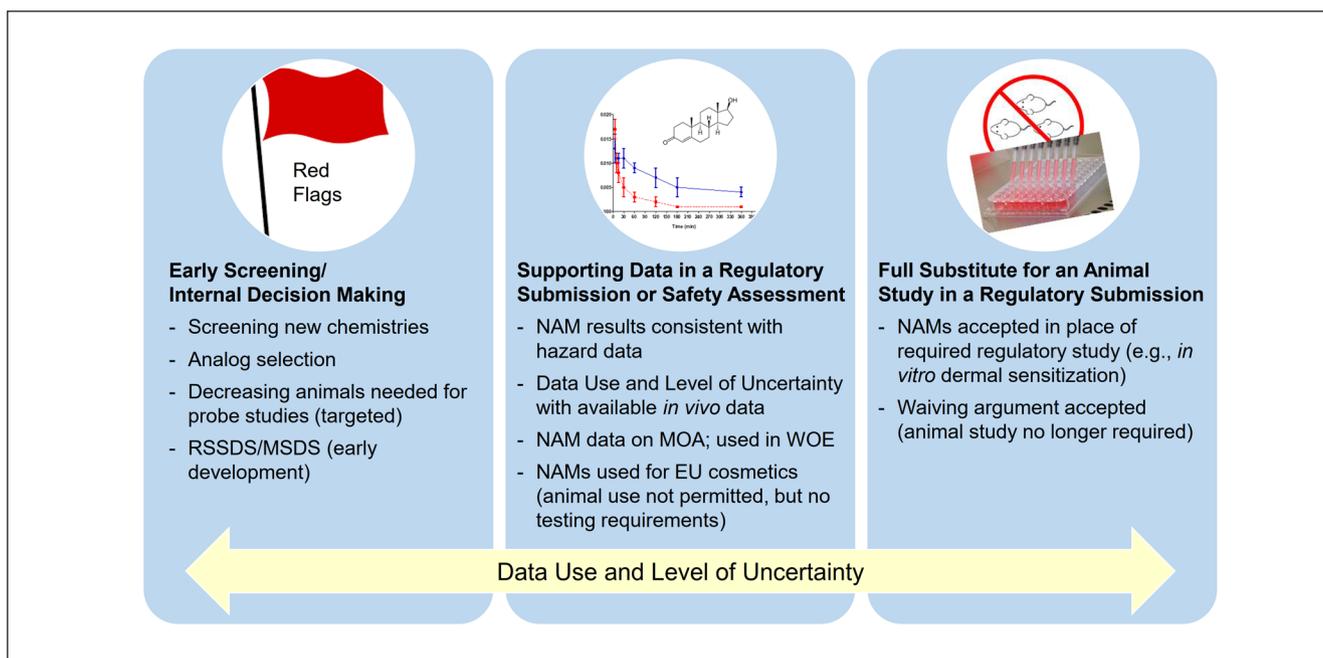


Fig. 1: Determining the appropriate number of animals to count towards animal savings depends on how NAM data are used and the level of certainty for decision-making

For example, for early screening or internal decision-making early in product development, studies typically involve rapid screening to identify “red flags” that would make a product unsuitable for development or to select among candidate compounds for a material with the most favorable hazard profile. In these cases, equivalent animal use values are often more speculative and, thus, should be conservative. To support data in a regulatory submission, assays are often used to address a specific data gap (targeted) or evaluate a broad swath of biological activity (non-targeted). In these cases, a unique study design or a partial guideline study may be used to set equivalent animal numbers. In the last scenario, full substitution for an animal study is achieved either by waiving arguments (e.g., read-across or exposure-based waiving) or NAMs that have achieved regulatory acceptance. In these cases, animal use reductions are easiest to calculate and are equivalent to the number of animals needed for the *in vivo* guideline study.

to evaluate dermal irritation, eye irritation/corrosion, and dermal sensitization potential, among other health effects. Internationally developed tools, including adverse outcome pathways¹ (OECD, 2015), integrated approaches to testing and assessment² (IATAs; OECD, 2016), and defined approaches to testing and assessment² (DAs; OECD, 2017) describe approaches to apply NAM data to regulatory uses.

As a result of this changing regulatory landscape, companies are developing and evaluating NAMs to meet data needs and to decrease animal testing. For this paper, NAMs include *in silico* computational models for structure-activity relationships or toxicokinetics, study waiving based on available information (e.g., read-across, physical-chemical properties, exposure-based waiving), or *in chemico* or *in vitro* models, all of which aid in internal decision-making regardless of the status of regulatory acceptance. In most companies, animal use policies emphasize the 3Rs (i.e., to replace, reduce and refine the use of animals in safety testing, whenever possible). Industry

and government agencies are dedicating resources towards the development and validation of NAMs. As a result, it has become increasingly important to develop metrics to track the implementation of NAMs and the decrease in animal use, as demonstrated in discussions among government bodies. The European Union has tracked the number of animals used for scientific purposes for decades, and, in their latest report, refined their tracking procedures to better identify where to focus the development and validation of alternatives (EC, 2020). In the U.S., the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) 2018 report “A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States” identified the need for metrics to prioritize activities and resources, monitor progress, and measure success of implementing NAMs (ICCVAM, 2018). The following year, the US Government Accountability Office (GAO) highlighted the need to identify metrics to monitor animal use in its recom-

¹ Collaborative Adverse Outcome Pathway Wiki (AOP-Wiki). Beta release in 2013 with subsequent updates. <https://aopwiki.org/> (accessed 17.07.2021)

² Integrated Approaches to Testing and Assessment and Defined Approach to Testing and Assessment. <https://ntp.niehs.nih.gov/whatwestudy/niceatm/integrated-testing-strategies/index.html> (accessed 17.07.2021)

**Tab. 1: Definitions of animals included in various animal use tracking programs**

Organization	Animal definition with inclusions/exclusions
Dow's "animal" definition is based on the American Association for Animal Laboratory Science (AALAS) Guide	<ul style="list-style-type: none"> • "Any vertebrate animal produced or used in research, teaching, or testing." • Animals ordered (preferred) or placed on study for toxicity testing • Animal number will include offspring (rats, fish, etc.) born during reproductive studies. • Animals will not include fetuses, embryos or other vertebrates prior to hatching, which is consistent with USDA^a (NIH, 2021) and Office of Laboratory Animal Welfare (OLAW) guidance^b. • Invertebrates are not included. • Animals monitored as part of field studies are excluded.
US EPA based on memo on animal use reductions (US EPA, 2019b)	<ul style="list-style-type: none"> • Focus on mammalian studies (i.e., per its footnote #1, the EPA memo applies to "whole animal or live mammalian studies and does not apply to use of mammalian cell cultures or human epidemiological studies").
European Union (EU, 2010)	<ul style="list-style-type: none"> • "Vertebrate animals used for experimental and other scientific purposes". • Includes fetal mammals in the last third of their development • Includes some invertebrates (e.g., cyclostomes, cephalopods) • Some flexibility for Member States to maintain national rules aimed at more extensive protection of animals

^a The NIH Guidelines for Preparing USDA Annual Reports and Assigning USDA Reporting Columns: "Number of Animals Bred, Conditioned or Held: Include all animals ...Used for breeding and their offspring, even if they were not used for research this reporting year. In the NIH IRP, offspring are counted at first cage change or experimental manipulation, whichever occurs first." (NIH, 2021); <https://oacu.oir.nih.gov/system/files/media/file/2021-02/a1-usda.pdf> (accessed 23.11.2021); ^b <https://olaw.nih.gov/faqs#/guidance/faqs?anchor=questionF> (accessed 23.11.2021) and Bartlett and Silk (2016) address and recognize fish as "animals" at "hatching" in accordance with Public Health Service Policy.

mentation to the National Institute of Environmental Health Sciences (US GAO, 2019). In response, an ICCVAM Metrics Workgroup published the report, "Measuring U.S. Federal Agency Progress toward Implementation of Alternative Methods in Toxicity Testing", recommending that each member agency develop metrics (ICCVAM, 2021).

To evaluate the utility of NAMs, provide a quantitative measure of accountability for resources spent on NAM development, and identify areas where their development is still needed, companies also can develop metrics to track toxicity testing – both numbers of animals used and numbers of animals not used due to NAMs – and thereby examine progress in the use of NAMs to provide information and reduce animal use. Thus, the goal of this paper is to present an approach developed by The Dow Chemical Company for estimating "animal savings or reductions in animal use" based on NAM use. The paper also lists points for companies or other organizations to consider when establishing their own tracking metrics to quantify progress towards the commonly stated goal to reduce and replace animal use. A central theme of this approach is that all NAM data that aid in decision-making have value (see Fig. 1). This paper proposes one approach to tracking reduced animal use, and it is anticipated that implementation at other organizations and input from other stakeholders will further improve tracking of animal use in toxicity testing to better illustrate the benefits of NAM use.

2 Methods

Animal definition for project scope

Any program designed to examine animal use must begin by defining what will be considered an "animal" in their tracking program. We adopted the definition of an animal from the American Association for Animal Laboratory Science (AALAS) Guide. According to the AALAS Guide, an animal is defined as "Any vertebrate animal produced or used in research, teaching, or testing". For purposes of determining the impact of NAMs, animal numbers will include animals ordered (preferred) or placed on study for toxicity testing for product safety, depending on available information. For internal studies, extra animals will be included when first brought into the lab but will not be counted again if placed on an alternate study. Dow's "animal" definition is further described in Table 1, which also highlights differences from the US Environmental Protection Agency (EPA) memo on animal use reductions (US EPA, 2019b), which focuses on mammals, and from the definition used in the European Union, which includes cephalopods, cyclostomes, and fetal mammals in the last third of their development (EU, 2010).

Establishing consistency in tracking animal use

To track progress toward reducing animal use, programs must develop their own guidance on how animal use will be measured.



First, it is important to establish an accurate baseline for the current number of animals used for toxicity studies as well as a protocol to consistently track animal use from year to year. Tracking in-house and external studies, both commissioned at contract research organizations (CROs) and funded (e.g., at universities), will allow for a better understanding of how study requirements shift from year to year and ensure that a reduction in internal animal use is not offset by an increase in external animal use. Ecotoxicology species (e.g., fish, tadpoles, birds) generated during stock colony breeding for study set-up will be tracked separately. Large numbers of fish and frogs may be used during study set-up, and, thus, the number of studies conducted in a given year could markedly impact overall animal numbers. Furthermore, CRO animal numbers typically do not report the number of animals used to set up ecotoxicity studies, thus making this number unavailable across all studies. Overall, it is beneficial to report separately on mammalian and non-mammalian ecotoxicological animal use numbers, species, and purpose of use (research and development, screening for internal decision-making, or regulatory requirements, as defined in Fig. 1) to provide clarity on animal use trends, because study set-up and/or reproduction studies with ecotoxicological species could dwarf numbers of mammalian animals used in toxicity testing, mask animal savings with NAMs, and conceal trends in animal use. Lastly, as business grows, regulatory requirements and animal use also may grow, so tracking business growth over time also may provide a useful perspective on animal use.

Identify goals

After a baseline number of animals used in testing is established, organizations can set goals to increase the development and uptake of animal-free testing approaches and reduce animal use. These goals will differ among industry, government agencies, and other organizations, and they may shift over time, for example, as companies' product portfolios change. For example, the US EPA had a goal to reduce mammalian studies by 30% by 2025 and eliminate all mammalian studies (except by Administrator exemption) although these dates are no longer mentioned in the 2021 version of the New Approach Methods Work Plan (US EPA, 2019a,b, 2021) and some companies adhere to a goal to conduct tests on animals only to comply with regulatory requirements³.

Methods to establish annual animal use numbers

The absolute number of both mammalian and aquatic animals can be tracked annually by monitoring animal orders or animals placed on study in-house. If the study includes a breeding phase, the number of offspring generated (or estimated if ecological species) also should be included. The number of animals used at CROs can be recorded using this same information. Animals used in studies as part of multi-company consortia or in studies funded at universities also should be tracked.

Sometimes animals are used to generate *in vitro* test systems, particularly those based on animal tissues. For example, some *in*

vitro toxicokinetic metabolism models (e.g., microsomes, S-9, or primary hepatocytes ordered from an external vendor) use animals for generation of the test system. In the current assessment, these NAMs are included as contributing to animal savings despite using animals for the generation of the test system. While not ideal, the rationale is that these isolated metabolic systems will allow for greater data generation per animal than *in vivo* work. In some cases, human tissues are available, which results in animal savings while increasing the relevance of these data for human risk assessment. If animals are used "in-house" or at CROs to generate tissues for NAM assessments, these animals are included in the "animal use" tally. While currently being discussed among stakeholders (van der Valk et al., 2004, 2010, 2018), animals are not counted for cell culture constituents (e.g., fetal bovine serum, fetal calf serum, or basement membrane matrices) or antibody generation (EURL ECVAM, 2020; Groff et al., 2020).

It is recognized that animal use will vary from year to year depending on regulatory programs, business markets/growth, and required study types. Therefore, the best metric over time may be a multi-year average of animal use, although information tracking the purpose of studies also will be useful. Furthermore, ecotoxicology studies and mammalian studies should be monitored separately due to fluctuations in study types from year to year, which can impact animal use numbers in a given year and provide important information to characterize animal use.

Inclusion of NAMs and other approaches

For this work, NAMs include *in silico* computational models to identify potential bioactivity/hazard or toxicokinetics, study waiving approaches, or *in chemico* or *in vitro* models (examples are described in the Results section and Tables 2-6 below). Furthermore, "intelligent designs" may allow researchers to consolidate multiple study endpoints into one study; thereby, negating the need to perform a separate, "stand-alone" study. If NAMs indicate a potential bioactivity of concern, "intelligent study designs" may allow researchers to collect additional endpoints or mode-of-action (MOA) data in an on-going or planned *in vivo* study, thereby avoiding a separate study. In these cases, animal savings from the avoided study would be included as equivalent animal savings.

Impact of NAMs on animal use

To calculate the NAM contribution to decreased animal use, there is a need to establish baseline rules on how to determine "equivalent animal savings" relative to *in vivo* study data. "Equivalent animal savings" is the estimated number of animals that would be used to generate equivalent information to what is provided by the NAM in question. This proposed approach recognizes that NAM data can provide information for a variety of decisions that have varying value (Fig. 1), which should be reflected in the "equivalent animal savings" number selected. Figure 2 shows a decision tree outlining some points to consider in assigning "animal savings" as described in the examples below:

³ e.g., https://www.unilever.com/Images/animal-testing-position-statement_tcm244-549936_en.pdf;
https://www.thecloroxcompany.com/wp-content/uploads/2019/10/Clorox-Animal-Testing-Policy-20190821_TCC.pdf

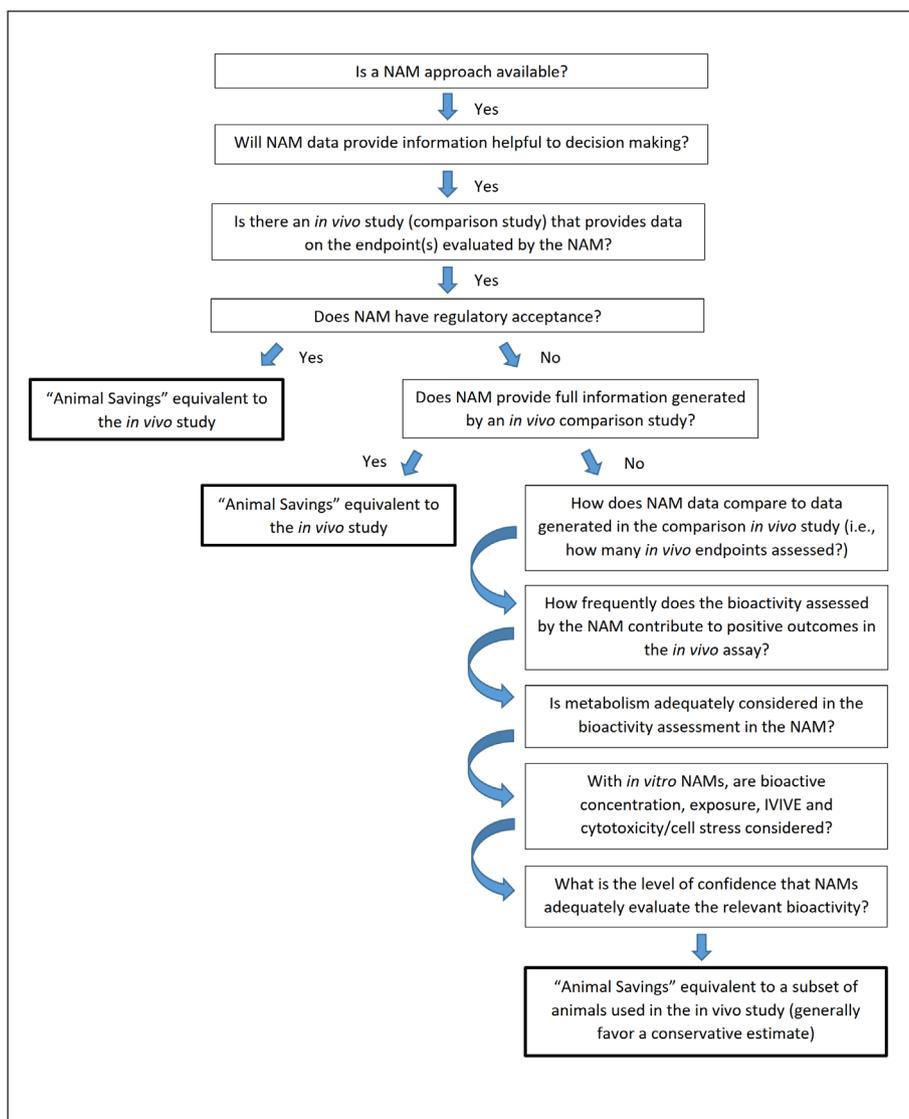


Fig. 2: A decision tree outlining points to consider when assigning “animal savings” due to NAM use

Some examples are provided in the text. When the NAM does not fully meet the information gleaned from an *in vivo* comparison study, estimates of “animal savings” are generally conservative. The level of confidence in the NAM outcome also is a critical element in selecting an “animal savings” value.

1. In some cases, NAM data have regulatory acceptance and alleviate the need to conduct an *in vivo* guideline study, in which case equivalent animal savings will be equal to the number of animals that would have been used in the conventional *in vivo* approach. For example, if two out of three *in vitro* assays allow for determination of dermal sensitization status in accordance with the OECD TG 497 “Defined Approaches on Skin Sensitization” (OECD, 2021), the “animal savings” could be up to 28, i.e., the number of animals used in a local lymph node assay (LLNA; OECD, 2010) with a positive control group.
2. NAM data may partially fulfill information generated by animal-based guideline studies, in which case the animal equivalent number for the NAM is a subset of the animal-based guideline study. For example, an androgen receptor (AR) transactivation assay with positive and negative controls can identify AR agonists and antagonists, whereas a Hershberger assay can detect androgen agonists, antagonists, and 5 α -reductase inhib-

itors. Furthermore, the Hershberger assay can evaluate metabolites, which are generally not evaluated in the AR transactivation assay. Thus, the AR reporter gene data are not fully equivalent to the 48 animals used in the Hershberger assay (OECD, 2009) but rather may be considered equivalent to 20% of the animals used in the full assay (e.g., 9.6 equivalent animal savings for each AR transactivation assay). Generally, the number selected should be conservative and reflect the degree of certainty in the results, considering issues like: a) frequency that the bioactivity assessed by the NAM contributes to positive outcomes in the *in vivo* assay in question (e.g., the reporter gene assay can detect AR agonists and antagonists, the bioactivities that drive many Hershberger “positive” results but cannot detect 5 α -reductase inhibitors); b) concentration, bioavailability and metabolism should be considered (e.g., highly metabolized compounds would have greater uncertainty as generally only the parent compound is tested in the AR trans-



activation assay); and c) confidence in the *in vitro* NAM, particularly if the assay is well characterized and the chemical is bioactive at doses below those causing cytotoxicity, cell stress and not at excessively high concentrations that are unlikely to be meaningful in an *in vivo* study.

- NAM data may provide information on bioactivity that is not accepted in a regulatory context but provides information for internal decision-making. In this case, the equivalent animal savings for the NAM is assigned on a case-by-case basis in line with the utility of the information. For example, if a chemical is identified as an aromatase inhibitor and can be detected by a QSAR model or *in vitro* assay, these NAMs could be used to screen analogs for aromatase inhibition to select candidate chemicals with a better hazard profile. In this case, this information has value equal to or less than an *in vivo* screening assay that examines aromatase activity depending on other endpoints assessed in the *in vivo* assay (e.g., count a 10% subset of animals from the pubertal female assay for each compound screened. The 10% “animal savings” is intentionally conservative, because while the pubertal female assay can identify aromatase inhibitors, it also can detect several other modes-of-action as well as evaluate bioactivity of metabolites).

Generally, in cases 2 and 3 above, a default number is assigned for NAM animal savings for each scenario (see Results section), but the animal savings number may be adjusted up or down depending on the specific scenario and degree of uncertainty. When this occurs, a note is included in tracking documents to explain why an adjustment was made. A conservative approach to animal savings through NAM use is preferred with the understanding that in the longer term, NAM animal savings should increase as more assays/batteries gain regulatory acceptance and can fully substitute for *in vivo* animal studies. Of course, the impact of NAMs on “percent reduction in animal use” may vary from year to year depending on regulatory requirements for *in vivo* studies.

One advantage of NAMs is that data generally can be obtained more quickly than in *in vivo* animal studies; however, a more protracted period may be required for regulatory agencies to determine the acceptability of NAM data in a regulatory context. Thus, when determining the contribution of NAMs to animal savings, there is a temporal component to information availability, regulatory review, and animal savings values. For example, if NAM data are used to fill data gaps (e.g., read-across argument or study waiver request) and, subsequently, regulators reject the read-across and require the *in vivo* animal study, the NAM may have been counted as animal savings at the time the read-across argument was posed, whereas later the animal numbers for the *in vivo* study would be counted towards animal use. These types of decisions are often separated by months or years, making this a necessary compromise in this approach. However, it can be anticipated that these types of reversals in animal use will diminish as NAMs gain more global regulatory acceptance. Furthermore, periodic retrospective analyses of read-across “successes and failures” may help to better position NAMs for regulatory acceptance in future submissions and improve animal savings.

Application of equivalent animal savings metrics in different scenarios

Determining animal savings from NAM applications is likely to be organization-specific, depending on internal practices, how data are used, and the level of certainty for decision-making (Fig. 1). For example, for internal decision-making early in product development (e.g., screening or prioritization), studies typically involve rapid screening to determine if a compound is suitable for development or to select an analog substance that has a better safety profile. Here, there are no specific animal numbers required in *in vivo* studies that were historically used to generate these data; thus, a conservative approach to equivalent animal savings is warranted, because: 1) there is greater uncertainty in NAM data collected early in product development as other contextual data are limited; and 2) typically, there are gaps in the NAM bioactivity assessment, and additional data collection will be needed as the substance moves further along the development process. When using NAMs to support data in a regulatory submission (e.g., dose-response or risk assessment), NAM data are used to supplement existing information, often to support or exclude a specific MOA. For example, for a TSCA submission, there was concern that a test compound would be metabolized to a teratogenic metabolite in humans despite a negative developmental toxicity study in rodents. To examine this, an *in vitro* comparative metabolism study was conducted to compare metabolism across species, including humans. This *in vitro* metabolism, which included related substances whose metabolism was known, was sufficient to support existing data and alleviate regulatory concerns. While there is no specific number of animals for this MOA study, a typical ADME study in one gender would require 4 animals/dose group to evaluate metabolite formation (i.e., determine C_{max}, then collect metabolites for identification); thus, it might be reasonable to assign equivalent animal savings of 4 animals for each group included in the *in vitro* metabolism study. With cosmetic ingredients, use of animal testing data generated after 2013 is banned in the European Union (EU, 2009); however, the requirements for safety testing of cosmetic ingredients are not strictly defined (although several options have been proposed, e.g., Baltazar et al., 2020). In such cases, some studies may be designed to examine generally required regulatory endpoints (e.g., genotoxicity) where equivalent animal savings may be easy to assign; however, these assessments also may include screening for a variety of bioactivities where assigning equivalent animal savings may be more speculative. Lastly, it is easier to assign equivalent animal savings values for read-across or exposure-based waiving arguments or *in vitro* assays (e.g., dermal sensitization) that have gained regulatory acceptance (i.e., equal to the number of animals needed to run the *in vivo* guideline study). As more NAMs gain regulatory acceptance, equivalent animal savings calculations will more accurately reflect animal savings.

Metrics on NAM reductions in animal use

After establishing equivalent animal savings for the information provided by each NAM in order to estimate the NAM impact on



animal use, each program can decide on the metrics of import for the use of NAMs. Some programs may wish to report absolute number of animal savings (i.e., equivalent animal savings) due to NAM use. Another option would be to report a “percent reduction in animal use” due to NAMs using the following two equations:

$$\sum AE = AE_{\text{vivo}} + AE_{\text{NAM}} \quad (\text{Eq.1}),$$

where $\sum AE$ is total animal equivalents, AE_{vivo} is number of animals used in *in vivo* studies, and AE_{NAM} is equivalent animal savings from NAM.

Then the percentage reduction in animal use, ΔAE (%), can be expressed as:

$$\Delta AE (\%) = 1 - \left(\frac{\sum AE - AE_{\text{NAM}}}{\sum AE} \right) * 100 \quad (\text{Eq.2}).$$

A third option might be to express the percentage of toxicity information that comes from NAMs, i.e. AE_{NAM} (%):

$$AE_{\text{NAM}} (\%) = \frac{AE_{\text{NAM}}}{\sum AE} * 100 \quad (\text{Eq.3}).$$

Depending on the goal, the calculations can be applied per chemical, per regulatory requirement or testing purpose, per species, or per endpoint or toxicity test.

3 Results

Our approach for tracking animal use reductions is shown in Tables 2-6. Generally, these tables show the endpoints assessed by a NAM, the corresponding *in vivo* test that provides similar information and the number of animals used in that study design, the proportion of equivalent animal savings relative to the *in vivo* study, a rationale to support the value assigned, and the default number of equivalent animal savings when employing the NAM approach. The equivalent animal savings numbers in these tables are default values and can be adjusted up or down depending on other available information (e.g., if modeling indicates that the test compound is highly metabolized, the number could be reduced). The rationale for adjustments to the default values can be captured in a spreadsheet or program tracking animal savings. We recognize that a single NAM may not fully mimic the situation in animals, as each NAM may be more limited in the number of endpoints evaluated and the ability to account for toxicokinetics (absorption, distribution, metabolism and elimination, ADME); therefore, we have tried to be conservative in our estimates of animal savings (e.g., often 10% of the number of animals used in the *in vivo* study). Other organizations may choose different values depending on how these data are used, tolerance of uncertainty, etc. Note that Tables 2-6 are not comprehensive but provide an overview of some common study types used by our laboratory.

Some examples of animal savings due to the use of *in silico* (computer-based) models are shown in Table 2. Dow has its own

cheminformatics group that applies publicly available QSAR models and builds its own models to predict toxicity. For example, models for acute oral toxicity have been under development for several years (e.g., Bhatarai et al., 2015; Wilson et al., 2018; Wijeyesakere et al., 2018, 2019, 2020). Our models can detect most of the potent MOAs for acute oral toxicity, but not all targets have been modeled. In addition, our models are generally conservative; while striving for high balanced accuracy is important, our models favor sensitivity over specificity to avoid missing false-negative chemicals to the extent possible. In addition, our *in silico* acute oral toxicity assessment considers both parent compound and potential metabolites (generally predicted via TIMES OASIS and GastroPlus™). Lastly, our acute oral toxicity model can predict GHS classification in most instances although these GHS designations have not achieved regulatory acceptance. Thus, the equivalent animal savings for an *in silico* assessment of acute oral toxicity was set at 30% of the *in vivo* acute oral toxicity study, resulting in a savings of 2.1 animals for each acute oral toxicity assessment. Equivalent animal savings for the application of some other *in silico* models also appear in Table 2. Note that these percentages can be adjusted depending on the certainty of the model predictions.

Table 3 lists examples of equivalent animal savings assigned for *in vitro* NAM assays and a brief rationale for the numbers selected. Some assays (e.g., acute endpoint assays) have received regulatory acceptance, and, thus, one assay or a combination of assays can fulfill a regulatory requirement. In these cases, equivalent animal savings were set to 100% of the *in vivo* study. In other cases, the *in vitro* NAM does not provide the full data set generated with an *in vivo* assay but provides information that can be used with other data in a weight-of-evidence approach for decision-making. For example, the *in vitro* steroidogenesis assay can identify altered androgen or estrogen synthesis in animals; however, the corresponding *in vivo* assay, the male or female pubertal assays, can identify numerous other bioactivities (e.g., ER/AR agonists, antagonists, thyroid active compounds). In addition, the pubertal assays can detect active metabolites. The notes under “Rationale” provide the bioactivities detected by the *in vivo* assay and limitations of the *in vitro* assay to detect these activities; these points were used to select an equivalent animal savings percentage for generating the NAM data. In the case of the Hershberger assay, many chemicals detected are AR antagonists (e.g., Luccio-Camelo and Prins, 2011), an activity that can be detected by the ARTA. However, some of these compounds require metabolism to generate active AR antagonists (Mansouri et al., 2020); thus, the value of the ARTA (without metabolic competency) was set at 20%. This proportion is a default value and may be increased based on other information (e.g., for poorly metabolized substances; use and regulatory acceptance of the COMPARA model; Mansouri et al., 2020).

Similarly, Table 4 identifies equivalent animal savings related to the use of *in silico* or *in vitro* ADME or bioaccumulation. Generally, *in vitro* models can provide valuable data on aspects like metabolism and metabolite identification (Dalvie et al., 2009). In some cases, ADME NAMs may not eliminate animal use due to the need for further information (e.g., time-course, distribution),



but may reduce animal numbers by decreasing the number of animals used to set dose levels, which also can be included in animal savings.

Animal use metrics also should include other aspects of the 3Rs for animal savings. This can include “intelligent study designs” that combine endpoints from different studies to increase the amount of information obtained from the same number of animals (Terry et al., 2014), or it may occur due to reduced numbers of animals used in probe study designs, staggered study starts to limit the number of dose levels needed, etc. Intelligent study designs have been a long-standing approach to reduce animal usage. Table 5 identifies some study types that can be integrated into repeat-dose studies to avoid conducting a separate study to assess these endpoints. Terry et al. (2014) describe an example of an agrochemical registration that successfully utilized intelligent designs for several required endpoints. Regulators have recognized the value of these approaches and have developed the extended one-generation reproductive toxicity study (EOGRTS, OECD 443; OECD, 2018a) as a design to examine endpoints in reproduction, endocrine and systemic toxicity, neurodevelopment, and the developing immune system all in one study, depending on which cohorts are included.

Sometimes studies can be waived for a variety of reasons (e.g., little or no exposure, not feasible to conduct a relevant study). One of the most applied approaches to waive *in vivo* studies is

by using read-across. Read-across extrapolates data from a related test substance or family of substances to predict the toxicity hazard of the compound in question and can be used to waive studies. Table 6 shows the number of animals used for a variety of *in vivo* studies to indicate potential animal savings with a successful waiving/read-across argument. There are numerous references offering guidance on how to prepare read-across evaluations (e.g., Ball et al., 2016). NAM and QSAR assessment often play a critical role in supporting read-across assessments. In cases where read-across is successful, this typically supplants the need to conduct the *in vivo* study and would result in animal savings numbers that are equivalent to the full *in vivo* study. In many cases, read-across arguments take months or years for review by regulatory agencies, and while registrants will be contacted if data are not acceptable, they may not be notified directly if these approaches are acceptable. Thus, animal savings for read-across generally fall into the year in which the read-across document is submitted. If regulators subsequently reject the “read-across” argument and the *in vivo* study is required, animals used in the *in vivo* study may be counted in a subsequent year. Given the temporal separation of these decisions and the changing regulatory landscape (e.g., acceptance of read-across, then subsequent request for data based on changing hazard concerns and/or changing data requirements), it may be difficult to retrospectively adjust animal savings numbers with any degree of accuracy.

Tab. 2: Animal use reductions due to the application of *in silico* (computer-based) NAMs to predict toxicity^a

Endpoint addressed by NAM	Corresponding <i>in vivo</i> test	No. animals <i>in vivo</i>	Animal savings using NAM	Rationale for percentage selected	No. animals saved by NAM use
Acute toxicity endpoints					
Acute oral toxicity	OECD ^b 425: Acute oral toxicity – up-and-down procedure; options: OECD 420 and 423	5-8 ^c ; 6-12 per Corvaro et al., 2016	30%	<ul style="list-style-type: none"> • Most potent MOAs identified, but not all relevant targets have been modeled • Predictions are conservative and route agnostic • Evaluation covers parent and predicted highly toxic metabolites • Predicts potential MOA and GHS class 	2.1
Acute dermal toxicity	OECD 402: Acute dermal toxicity – fixed dose procedure	4-9 ^d ; avg = 7; full test 50 per Corvaro et al., 2016	30%	<ul style="list-style-type: none"> • Similar predictions to acute oral toxicity are used and generally considered conservative 	2.1
Acute inhalation toxicity	OECD 403: Acute toxic class; other options: OECD 436 and 433	40-42 ^e	10%	<ul style="list-style-type: none"> • Not all relevant targets have been modeled • Predictions are conservative and route agnostic • Evaluation covers parent and predicted metabolites • Predicts potential MOA but not GHS class 	4
Acute dermal irritation/corrosion	OECD 404: Acute dermal irritation/corrosion	2-3	10%	<ul style="list-style-type: none"> • Use read-across and mechanistic profiling • Evaluation covers parent • Predict pos. or neg., but not GHS class (unless using read-across) • If needed, <i>in vitro</i> follow-up to address uncertainty 	0.3



Endpoint addressed by NAM	Corresponding <i>in vivo</i> test	No. animals <i>in vivo</i>	Animal savings using NAM	Rationale for percentage selected	No. animals saved by NAM use
Eye irritation	OECD 405: Acute eye irritation/corrosion	1-3	10%	<ul style="list-style-type: none"> • Use read-across and mechanistic profiling • Evaluation covers parent • Predict pos. or neg., but not GHS class (unless using read-across) • If needed, <i>in vitro</i> follow-up to address uncertainty 	0.3
Skin sensitization	OECD 429 or 442A: LLNA; option: 406	28 ^f	30%	<ul style="list-style-type: none"> • Most relevant bioactivities (e.g., facile reactivity) included in models/profilers • Predict pos. or neg., but not GHS class (unless using read-across) 	8.4
Filling data gaps for Research Sample Safety Data Sheets (RSSDS) or Material Safety Data Sheets (MSDS)	<ul style="list-style-type: none"> • OECD 425: Acute oral • OECD 404: Acute dermal irritation/corrosion • OECD 405: Eye irritation • OECD 429: LLNA 	41 ^g	30%	<ul style="list-style-type: none"> • Not all relevant targets have been modeled • Screening novel chemical domains • Predictions are conservative and route agnostic; reasonable confidence • Evaluation covers parent and predicted metabolites • Predicts potential MOA but not GHS class for all endpoints (acute oral only) 	12.3
Ecotoxicological QSAR predictions					
Acute fish toxicity	OECD 203: Fish, acute toxicity test	42 ^h -49	30%	<ul style="list-style-type: none"> • Models used for PMN and REACH submissions if chemical is in domain • Long-term use with reasonable understanding of reliability 	14.7
Chronic fish toxicity	OECD 210: Fish, early life stage toxicity test	560 ⁱ (840 ^j)	30%		168
Bioaccumulation (BCF)	OECD 305: Bioaccumulation in fish	251 ^k	30%		75.3

^a Publicly available and internally developed databases with examples of model applications (e.g., Wilson et al., 2018; Wijeyesakere et al., 2018, 2019, 2020; Krieger et al., submitted); *in silico* evaluations also include an evaluation of potential metabolites identified through TIMES OASIS and/or GastroPlus™. ^b OECD Guidelines for the Testing of Chemicals: <https://www.oecd.org/chemicalsafety/testing/oecdguidelinesforthetestingofchemicals.htm>. ^c Estimated average number depending on successful 2000 mg/kg/day limit dose approach (5 animals) or up-and-down main study estimate where the stopping rule is satisfied using 4-6 animals after test reversal and assuming that the reversal occurs at the second dose level tested. ^d 2-3 animals for range-finding study and 2-6 animals in the main study. ^e 6 animals (3/sex) in the sighting study (assumes 2 concentrations); 10 animals (5/sex) at 3 concentrations in the main study. Minimum animal use for limit concentration requiring 6 animals (3/sex). With *C x t* approach, 2 animals at 4 concentrations at 5 exposure durations = 40 animals. A concurrent control is generally not required unless data on vehicle control is lacking. ^f 4 animals/dose with 3 dose levels plus a negative (vehicle) control group and a positive control group (20 animals) plus 1-2 animals/group for preliminary irritation assessment (e.g., control, 3 concentrations) = 28. ^g Includes 7 (acute oral) + 3 (dermal irritation) + 3 (eye irritation) + 28 (LLNA) = 41 animals. ^h Definitive test: 7 fish/concentration with minimum of 5 concentrations plus 1 dilution water control and, if applicable, 1 vehicle control (although limit test can be run with 14 fish); does not include fish needed for rangefinder if required (~18-30 fish). ⁱ Test requires at least 80 eggs per concentration (20 eggs/replicate) with a minimum of 5 test concentrations plus 1 dilution water control and, if applicable, 1 vehicle control; does not include fish needed for rangefinder if required (~70-210 fish). ^j Many labs conduct tests with up to 120 eggs per concentration (30 eggs/replicate) and thin post-hatch to 80 larvae per concentration (20 larvae/replicate) with a minimum of 5 test concentrations plus 1 dilution water control and, if applicable, 1 vehicle control. ^k Full aqueous exposure: 4 fish per sampling time point conducted with two concentrations plus 1 dilution water or vehicle control group sampled at least 5 times during the uptake phase and 4 times during the depuration phase (12 fish at 9 time points = 109 fish). Dietary bioaccumulation test uses additional fish (5-10 fish at each time point with 2 timepoints during the uptake/assimilation phase and 4-6 sampling times during the depuration phase with 1 test concentration and a control group = 160 fish total). However, this value does not include fish that may be collected during study from the control and each concentration for lipid analyses (~12 fish each); for parent substance/metabolite analyses via HPLC (~36 fish each) and, if applicable, for metabolite identification (~30 fish each) for the control and each treatment (~102 fish total). Furthermore, additional fish (108 fish) may be necessary should the study duration be extended for the maximum 60-day exposure (60 fish) and/or maximum 56-day depuration phase (48 fish) to reach steady-state and/or adequate reduction in body burden of the test substance, respectively (108 total). A preliminary test (n ≥ 40 fish) can aid in determination of accumulation rate and help to better define study design. Therefore, estimated 40 + 109 + 102 = 251 fish.

Tab. 3: Animal use reductions due to the application of *in vitro* NAMs to predict toxicity^a

Endpoint addressed by NAM	Corresponding <i>in vivo</i> test	No. of animals <i>in vivo</i>	Animal savings using NAM	Rationale for percentage selected	No. animals saved by NAM use
Acute endpoints					
3T3 neutral red uptake (NRU) (OECD ^a 432) for phototoxicity/ cytotoxicity	No formal OECD phototoxicity/ cytotoxicity test guideline; estimate based on other dermal studies	2-3	100%	Equivalent to phototoxicity screen	2
Reactive oxygen species (ROS) for photo-reactivity (OECD 495)					
Transcutaneous electrical resistance test (OECD 430), membrane barrier test (OECD 435) or reconstructed human epidermis (RhE) test (OECD 431) for skin corrosion	OECD 404: Acute dermal irritation/ corrosion	2-3	100%	Equivalent to skin irritation screen	3
Reconstructed human epidermis (RhE) test (OECD 439) for <i>in vitro</i> skin irritation					
Fluorescein leakage (OECD 460) or Statens Serum Institut rabbit cornea (SIRC) cells for eye corrosives or severe irritants (OECD 491)	OECD 405: Acute eye irritation/ corrosion	1-3 (1 if corrosive)	100%	Equivalent to eye irritation screen	3
Reconstructed human cornea-like epithelium (RhCE) test for eye irritation					
Bovine corneal opacity test (BCOP; OECD 437), isolated chicken eye test (OECD 438), or Vitrigel (OECD 494) for eye irritation					
ARE-Nrf2 luciferase KeratinoSens™ or ARE-Nrf2 luciferase LuSens test (OECD 442D)	OECD 429 or 442A: LLNA; option: 406	28 ^b	33-50%	2-3 assays are needed to equal <i>in vivo</i> assessment; if both assays agree, then 50% animal savings for each assay. If all three assays are needed, 33% savings for each assay. Total animal savings equals 28 as these assays have regulatory acceptance and a combination can replace the LLNA.	28
Direct peptide reactivity assay (DPRA) or amino acid derivative reactivity assay (ADRA) (OECD 442C)			33-50%		
Human cell line activation test (h-CLAT), U937 cell line activation test (U-SENS™), or interleukin-8 reporter gene assay (IL-8 Luc assay) (OECD 442E)			33-50%		
Endocrine endpoints					
ER transactivation assay (ERTA) (OECD 455; binding only in OECD 493)	OECD 440: Uterotrophic bioassay	18 ^c	50-100%	<ul style="list-style-type: none"> • Uterotrophic can measure agonists and antagonists (test guideline agonists only) • ERTA focus on parent compound, not metabolites • When coupled with ER Expert System (CERAPP), may fulfill requirements to screen for estrogenicity 	9-18
AR transactivation assay (ARTA) (OECD 458)	OECD 441: Hershberger assay	48 ^d	20%	<ul style="list-style-type: none"> • Hershberger can measure AR agonist and antagonists, 5-alpha-reductase inhibitors (5αR I); AR cannot detect 5αR I • ARTA focus on parent compound, not metabolites 	9.6
Aromatase assay (OPPTS 890.1200)	OPPTS 890.1450: Female pubertal assay	60	10%	<ul style="list-style-type: none"> • Female pubertal also can detect steroidogenesis inhibitors, E2, anti-E2, androgens, thyroid; aromatase assay only detects limited subset (i.e., only one 	6



Endpoint addressed by NAM	Corresponding <i>in vivo</i> test	No. of animals <i>in vivo</i>	Animal savings using NAM	Rationale for percentage selected	No. animals saved by NAM use
				enzyme in pathway, although one that is commonly affected by chemicals) • Aromatase assessment primarily of parent compound	
H295R steroidogenesis assay (OECD 456)	OPPTS 890.1450 or 890.1500: Female or male pubertal assay	60	10%	• Pubertal assays also can detect E2, anti-E2, androgen, anti-androgens, steroidogenesis inhibitors and several possible thyroid MOAs (enzyme induction and T4 clearance, TPO, carrier protein binding displacement, deiodinase inhibition, NIS inhibition) • Steroidogenesis assessment primarily of parent compound	6
TPO assay	OPPTS 890.1450 or 890.1500: Female or male pubertal assay	60	5	• Pubertal assays also can detect E2, anti-E2, androgen, anti-androgens, steroidogenesis inhibitors and several possible thyroid MOAs (enzyme induction and T4 clearance, TPO, carrier protein binding displacement, deiodinase inhibition, NIS inhibition) • TPO assessment primarily of parent compound	3
Genotoxicity endpoints					
Gene mutation: Ames assay (bacterial OECD 471) or mammalian <i>Hgprt/Xrpt</i> (OECD 476) or Tk (OECD 490)	OECD 488: Transgenic rodent, OECD TBD: Pig-A, or OECD 489: Comet	≥ 40	0-10%	• Assay examines parent and metabolites • Apply animal savings depending on application and whether <i>in vivo</i> tests will be required as follow-up ^e	4
<i>In vitro</i> micronucleus test (MNT) (OECD 487) or <i>in vitro</i> mammalian chromosomal aberration test (OECD 473)	OECD 474: <i>In vivo</i> MNT or Comet	≥ 40	0-10%		4
Ecotoxicological endpoints					
<i>In vitro</i> mitochondrial inhibition	OECD 203: Fish, acute toxicity test	42 ^f -49	10%	• Only useful for compounds positive for mitochondrial inhibition at low concentrations (no savings for negative/weak compounds) • ID GHS Cat. 1 for fish and daphnia • Relationship between mitochondrial toxicity and acute fish toxicity described in Bhatarai et al. (2015)	4.9

^a OECD Guidelines for the Testing of Chemicals: <https://www.oecd.org/chemicalsafety/testing/oecdguidelinesforthetestingofchemicals.htm>.

^b 4 animals/dose with 3 dose levels plus a negative (vehicle) control group and a positive control group plus 1-2 animals/group for preliminary irritation assessment (e.g., control, 3 concentrations) = 28. ^c Minimum of two dose groups and a control (6 animals/each) and assumes a separate 'stand-alone' positive control study for 17 α -ethynyl estradiol (not included) and no dose-range finding study. ^d 6 animals/group with a minimum of two dose groups and a positive (testosterone) and negative (vehicle) control for androgenicity and three dose groups plus a positive (testosterone + flutamide) and negative (testosterone, already included) control for antiandrogenicity (8 groups).

^e Actual animal usage will depend on geographical and application-specific requirements (e.g., candidate selection, registration, production volume (geography-dependent)) and results of *in vitro* battery. In some regulatory programs, an *in vivo* test for genotoxicity is required (no animal savings). If a positive *in vitro* genotoxicity result occurs during registration in some programs, an *in vivo* follow-up study is required, and animal savings should not be counted. However, for other applications (e.g., if a positive *in vitro* genotoxicity result prevents a candidate chemistry from moving forward in development), animal savings may be counted. ^f Definitive test: 7 fish/concentration with minimum of 5 concentrations plus 1 dilution water control and, if applicable, 1 vehicle control (although limit test can be run with 14 fish); does not include fish needed for rangefinder if required (~18-30 fish). TBD, to be determined as the test guideline is not yet finalized.


Tab. 4: Animal use reductions due to *in silico/in vitro* metabolism or bioaccumulation

Endpoint addressed by NAM	Corresponding <i>in vivo</i> test	No. of animals <i>in vivo</i>	Animal savings using NAM	Rationale for percentage selected	No. of animals saved by NAM use
<i>In vitro</i> comparative metabolism (IVCM) (mouse, rat, rabbit, dog, human)	OECD ^a 417: Toxicokinetics (absorption, distribution, metabolism, and elimination; ADME) ^b with inclusion of multiple species for cross-species comparison	Minimum 4 rats, 8 mice and 2 rabbits for cross-species comparison; 2 dogs for absorption, metabolism, and elimination (AME) = 16	50% (max. savings = 8 if all species included)	<ul style="list-style-type: none"> Probe AME covers 3 species (rat, mouse, rabbit) <i>In vivo</i> ADME study also tracks absorption, distribution, time course and elimination of radiotracer Only metabolism covered <i>in vitro</i> IVCM has 5 species (pools are n=3 individuals/pool) 	8 (if all species evaluated)
OECD 428: Skin absorption	OECD 427: Skin absorption ^c	12	50%	<ul style="list-style-type: none"> Measure diffusion of chemicals into/ across skin (non-viable skin model) or measure diffusion and metabolism (fresh, metabolically active skin) Human skin can be used <i>In vivo</i> study tracks distribution and elimination of radiotracer 	6
<i>In silico</i> GastroPlus modeling (REACH ADME) ^d	OECD 417: Toxicokinetics (ADME) ^b	8 rats/route x 3 routes = 24 rats; 4 rats/route = 12 rats if no gender difference	1 animal/route; maximum savings = 3	<ul style="list-style-type: none"> REACH dossier support <i>In vivo</i> ADME study also tracks absorption, distribution, time course and elimination of radiotracer GastroPlus – Route-specific AUC, Cmax, Tmax, time-course and metabolite predictions; half-life (25% savings without gender evaluation) 	3 (or 6 if there are gender differences predicted)
<i>In silico</i> GastroPlus modeling for IVIVE			25% of 4 rats	<ul style="list-style-type: none"> Explore <i>in vivo</i> concentration relative to external dose (one route) Minimum 2 animals/dose with 2 doses to understand <i>in vivo</i> dosimetry 	1
<i>In vitro</i> fish hepatocytes/ S9 for bioaccumulation	OECD 305: Bioaccumulation in fish	108 ^f (265)	30%	<ul style="list-style-type: none"> Strengthens weight-of-evidence for BCF derived from QSAR (If chemical is in QSAR domain and metabolized in this <i>in vitro</i> assay, this may be sufficient for bioaccumulation endpoint) Assay can be used to screen candidate chemistries Not accepted for bioaccumulation in all geographies 	200

^a OECD Test Guidelines are located at: <https://www.oecd.org/chemicalsafety/testing/oecdguidelinesforthetestingofchemicals.htm>. Relevant guidelines in Sections 2: Effects on Biotic Systems and Section 4: Health Effects. ^b 4 animals/dose in a single dose pilot study and two dose levels in the main study (12 animals) for each route. ^c 4 animals/dose group/timepoint with test material administered at 1 dose and samples collected at 3 time points (at the end of exposure and two subsequent occasions); thus, 12 total animals. ^d Validation of GastroPlus™ reported (Zhang et al., 2018); potential metabolites also identified through TIMES OASIS and/or GastroPlus™. ^e Full aqueous exposure: 4 fish per sampling time point conducted with two concentrations plus 1 dilution water or vehicle control group sampled at least 5 times during the uptake phase and 4 times during the depuration phase. Dietary bioaccumulation test uses additional fish, requiring sampling of 5-10 fish at each time point with 2 timepoints during the uptake/assimilation phase and 4-6 sampling times during the depuration phase with 1 test concentration and a control group (160 fish total). ^f Full aqueous exposure: 4 fish per sampling time point conducted with two concentrations plus 1 dilution water or vehicle control group sampled at least 5 times during the uptake phase and 4 times during the depuration phase. Dietary bioaccumulation test uses additional fish, requiring sampling of 5-10 fish at each time point with 2 timepoints during the uptake/assimilation phase and 4-6 sampling times during the depuration phase with 1 test concentration and a control group (160 fish total). However, this value does not include fish that may be collected during study from the control and each concentration for lipid analyses (~12 fish each); for parent substance/metabolites analyses via HPLC (~36 fish each) and, if applicable, for metabolite identification (~30 fish each) for the control and each treatment (~102 fish total). Furthermore, additional fish (108 fish) may be necessary should the study duration be extended for the maximum 60-day exposure (60 fish) and/or maximum 56-day depuration phase (48 fish) to reach steady-state and/or adequate reduction in body burden of the test substance, respectively (108 total). Thus, 265 was selected as an intermediate number of fish used for this study.

Tab. 5: Animal use reductions due to intelligent study designs^a

Endpoint integrated into another repeat-dose animal study	No. of animals <i>in vivo</i>	Animal savings using intelligent designs	Rationale for percentage selected	No. of animals saved by intelligent design
OECD TBD: Pig-A gene mutation assay	30 ^b (24 excluding positive controls)	80%	• Integrated in OECD 407 (28-day), 408 (90-day) or OECD 421/422 (repro screens)	24
OECD 474: <i>In vivo</i> micronucleus test (OECD 474)	25 ^c	100%	• Integrated in OECD 407 (28-day), 408 (90-day) or OECD 421/422 (repro screens)	25
T-cell dependent antibody response (TDAR)	45 ^d	100%	• Integrated in OECD 407 (28-day), 408 (90-day) or OECD 443 (EOGRS) • Integrated in a 90-d study (Terry et al., 2014, as described in Ladics et al., 1995)	45
OECD 424: Neurotoxicity	80	100%	• Similar to OECD 424 • Integrated in a 90-d study (Terry et al., 2014)	80
OECD 417: TK/metabolism (blood and urine collection during repeat-dose studies)	8 rats/route x 3 routes = 24 rats; 4 rats/route = 12 rats if no gender difference	50%	• Integrated in a 90-d study (Terry et al., 2014) to ID blood levels and excretion at steady-state • <i>In vivo</i> ADME study tracks absorption, distribution, time course and elimination of radiotracer	6-12

^a OECD Test Guidelines are located at: <https://www.oecd.org/chemicalsafety/testing/oecdguidelinesforthetestingofchemicals.htm>
Relevant guidelines in Sections 2: Effects on Biotic Systems and Section 4: Health Effects; TBD, to be determined as the test guideline is not yet finalized. ^b n = 6/dose group x 5 groups (3 treated groups + positive and negative control groups), but laboratories may run a separate positive control or choose not to include an *in vivo* positive control in each study (in both cases, animal savings = 24); numbers may differ in the final test guideline. ^c n = 5/dose group x 5 groups (3 treated groups + positive and negative control groups); If there is a difference in sensitivity, 50 animals (5/sex/dose) are required but this is atypical. ^d Assumes 10/dose group x 4 groups (3 treated groups + control) plus 5 positive control animals.

Tab. 6: Some examples of 100% animal savings if regulatory toxicity studies are waived

Test	Name	Adults	Fetuses ^a /Pups	Total
Acute systemic toxicity (oral, dermal, and inhalation)				
OECD ^b 425, 402, 403	Waiving based on physico-chemical properties of the test chemical (not feasible, relevant or little/no exposure potential by a specific route, existing acute oral toxicity data) or animal welfare considerations ^c	Savings dependent on route(s) for which waiving is applicable 5 or 8 ^d (oral) + 50 ^e (dermal) + 40 ^{e,f} (inhalation) = 95 or 98	NA	95 or 98 if all routes
	Read-across using existing hazard information ^c			
	GHS additivity formula for mixtures ^b (e.g., Corvaro et al., 2016)			
Repeat-dose toxicity				
OECD 407	28-day repeat-dose toxicity study	60 ^g	NA	60
OECD 408	90-day repeat-dose toxicity study	100 ^h		120
OECD 453	Combined chronic/carcinogenicity study	560 ⁱ		560
Developmental and reproductive toxicity				
OECD 421/422	Reproductive/developmental toxicity screening study with and without repeated-dose toxicity study	80 ^j	520 ^l	600
OECD 443	Extended one-generation reproductive toxicity study (EOGRS)	200 ^k	1040 ^l	1240



Test	Name	Adults	Fetuses ^a /Pups	Total
OECD 416	Two-generation reproductive toxicity study	184 ^m	2080 ^l	2264
OECD 414	Developmental toxicity study (rat or rabbit)	92 to 100 ^m	(1040 rat or 640 rabbit ^l)	92 to 100
OECD 426	Developmental neurotoxicity study	92 ^m	1040 ^l	1132
Endocrine toxicity				
OECD 440	Uterotrophic assay	18 ⁿ	NA	18
OECD 441	Hershberger assay	48 ^o		48
OPPTS 890.1500	Male pubertal assay	60		60
OPPTS 890.1450	Female pubertal assay	60		60
US EPA Guidance	Comparative thyroid study	184 ^p	(1040 ^{a,p}) + 1040 ^p	1224

^a Animal savings programs may vary in their decision to include fetuses in animal counts; not included in the current paper per our “animal” definition. ^b OECD Test Guidelines are located at: <https://www.oecd.org/chemicalsafety/testing/oecdguidelinesforthetestingofchemicals.htm>. Relevant guidelines in Section 2: Effects on Biotic Systems and Section 4: Health Effects. ^c OECD Guidance document no. 237 available for waiving acute toxicity tests in mammals. ^d Estimated average number depending on successful 2000 mg/kg/day limit dose approach (5 animals) or up-and-down main study estimate where the stopping rule is satisfied using 4–6 animals after test reversal and assuming that the reversal occurs at the second dose level tested. ^e Limit test uses 10 animals and thus, for non-toxic substances, animal use numbers should be adjusted accordingly. ^f 6 animals (3/sex) in the probe study (assumes 2 concentrations); 10 animals (5/sex) at 3 concentrations in the main study. Minimum animal use for limit concentration requiring 6 animals (3/sex). With C x t approach, 2 animals at 4 concentrations at 5 exposure durations = 40 animals. A concurrent control is generally not required unless data on vehicle control is lacking. ^g 5/sex/dose group with 4 dose levels plus 5/sex/dose group in the control and high-dose group to examine reversibility. ^h 10/sex/dose group with 4 dose levels plus 5/sex/dose group in the control and high-dose group to examine reversibility. ⁱ Assumes 70 animals/sex/dose group with 4 dose levels to allow removal of 10 animals/sex/dose group at 1 year for chronic evaluation and 60 animals/sex/dose to meet survival requirements for a valid study. ^j 10/sex/dose group to produce 8 pregnant females/dose group. ^k Given the size and complexity of the EOGRTS, this assumes 25/sex/dose group to produce 20 pregnant females/dose group. ^l Assumes an average of 13 pups/litter (rat) or 8 pups/litter (rabbit). ^m For rats, assumes a starting number of 23/sex/dose group or 23 time-mated dams/dose group to produce 20 pregnant females/dose group. For rabbits, assumes a starting number of 25 time-mated does/dose group to produce 20 pregnant females/dose group. ⁿ Minimum of two dose groups and a control (6 animals/each) and assumes a separate “stand alone” positive control study for 17 α -ethinyl estradiol (not included) and no dose-range finding study. ^o 6 animals/group with a minimum of two dose groups and a positive (testosterone) and negative (vehicle) control for androgenicity and three dose groups plus a positive (testosterone + flutamide) and negative (testosterone, already included) control for antiandrogenicity (8 groups). ^p Assumes a starting number of 46 dams/dose group to produce 40 pregnant females/dose group with each pregnancy having 13 offspring; therefore, 1040 fetuses for prenatal assessments and 1040 for postnatal assessments in the offspring. NA, not applicable

4 Discussion

This paper describes one approach that can be used to track NAM contribution to reducing animal use. The foundation of our metrics is that NAM data have significant value and that it would have taken animals to provide this information in the absence of NAMs. Even smaller, interim decisions (e.g., moving a compound forward in development) can benefit from the use of these predictive tools. Basically, the request and subsequent conduct of a predictive toxicology assessment indicates that the information is needed and, therefore, has value for decision-making. For example, *in vitro* and *in silico* approaches can be used to select candidate chemicals or to inform further testing by refining study designs and reducing the use of large numbers of animals. *In silico* and *in vitro* approaches also can improve dose selection, requiring fewer animals for dose-finding studies. Thus, animal savings will be included for any NAMs that provide useful information for decision-making, in addition to methods that directly replace the use of animals.

In determining the proportion of animal savings that is assigned for a given NAM, it is important to identify how the NAM data fulfill a testing/data need compared to an *in vivo* study conducted for a similar purpose. NAM equivalent animal savings should be adjusted to account for the scope of *in vitro* data relative to *in vivo* data. This approach assumes that the NAMs employed are “fit for purpose”, having been evaluated for performance, sensitivity, robustness, and domain of applicability. Notably, regulatory acceptance is not required for a NAM to have value (Archibald et al., 2015).

Implementing NAMs and tracking animal savings provide a positive return on investment for companies and other organizations. NAMs allow for more rapid data generation, sometimes at lower cost (e.g., Meigs et al., 2018) and, in some cases, with greater human relevance (Clippinger et al., 2021). In addition, the procedures for product safety assessments are undergoing an evolution, requiring laboratories to integrate multiple data streams in IATA. Our generic IATA template is shown in Figure 3. This IATA approach starts with cheminformatics

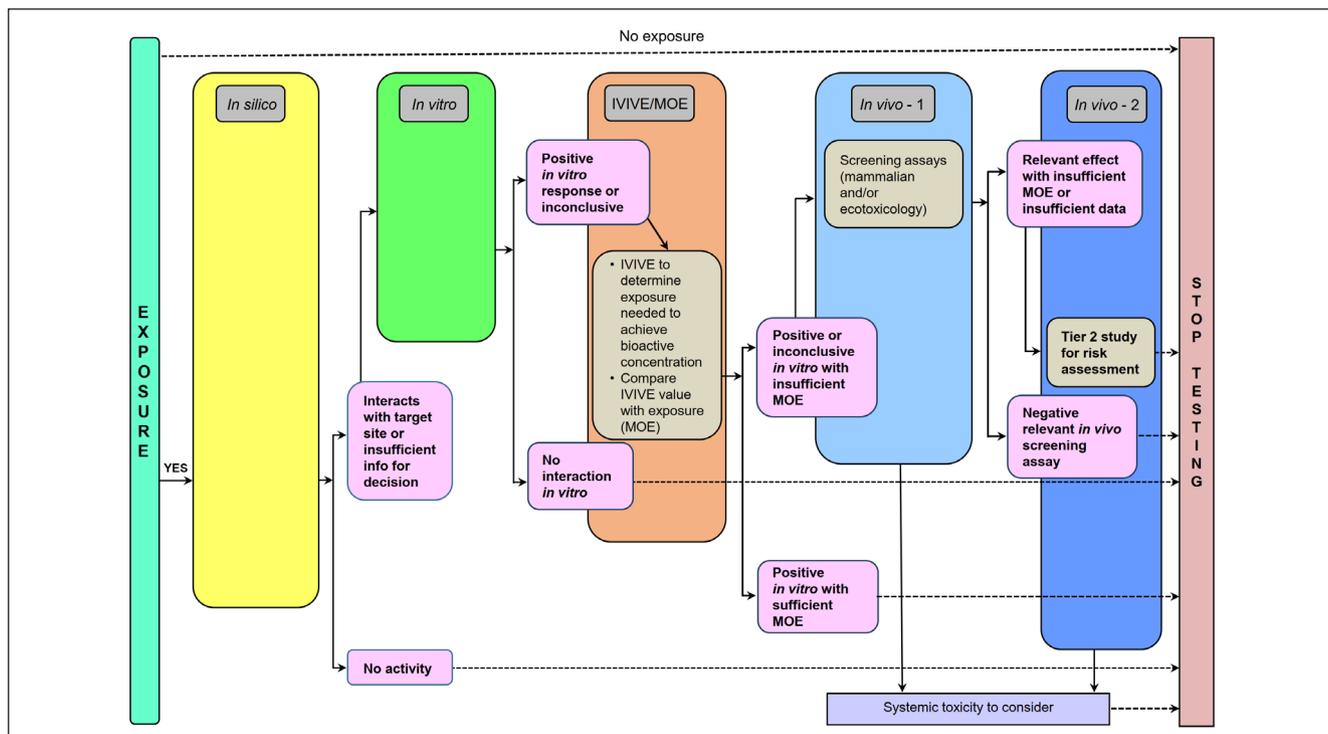


Fig. 3: Our generic template for an integrated approach to testing and assessment (IATA) of chemicals

As shown, NAMs are important elements of this approach, starting on the left with *in silico* approaches (cheminformatics, QSARs, read-across), quantitative *in vitro* assessments, and quantitative *in vitro*-to-*in vivo* (QIVIVE) evaluations to determine the relevance of positive *in vitro* findings. This IATA approach can reduce the need for *in vivo* studies or refine *in vivo* study designs to limit the number of animals used, depending on final data requirements. MOE, margin of exposure

ics, QSARs and read-across to identify potential toxicity (e.g., Luechtefeld et al., 2018). *In silico* predictions can be further evaluated using *in chemico* or *in vitro* approaches along with quantitative *in vitro*-to-*in vivo* extrapolation (QIVIVE) to provide relevance for any positive results. Together, these data can be used on a case-by-case basis to fill risk assessment data gaps, bridge an animal dataset to the human situation, or prioritize follow-up *in vivo* studies (EFSA, 2014). In addition, the IATA approach provides a framework for organizing NAM data for internal decision-making, providing insights on data gaps and confidence.

There are challenges to implementing NAMs and tracking animal savings metrics too. For example, there may be insufficient coverage of biological space with available NAMs or a need to include/optimize metabolic competence of *in vitro* assays to maximize their human relevance. This results in reductions to the animal savings metric for NAMs, which is application-dependent and may vary depending on other available information. Other aspects of tracking animal use further complicate measurement of animal savings. For example, large numbers of fish may be bred during mating as a study phase or for study set-up. This number of animals may minimize important reductions that are occurring elsewhere in a program implementing NAMs; thus, aquatic species should be tracked separately. In addition, studies conducted through a consortium may be counted by more than one company; therefore, it may be useful to note this for any test-

ing conducted for these purposes to avoid duplicating data if it is combined with other companies.

Another challenge in defining a tracking system is identifying NAMs that may not replace animal use but increase knowledge of human health effects. For example, an *in vitro* method to assess respiratory sensitization is currently being developed and may provide a valuable contribution to evaluating a substance's effect on human health that is not currently able to be assessed using *in vivo* methods (Chary et al., 2018, 2019). While such tests are not accounted for within the tracking system identified here, they are important to monitor progress.

This paper provides some insights into the approaches used at Dow to track reduction in animal use. This approach can be adopted by other organizations with modification as needed depending on how data are used, tolerance for uncertainty in decision-making, etc. In any event, decisions on equivalent animal savings for NAMs should be clearly documented to ensure consistent application from year to year. Work done at CROs also should be included. Averaging animal savings counts over multiple years may be useful to avoid variability in animal use due to changes in regulatory requirements (e.g., requirement for numerous reproductive studies may increase animal use relative to other years) while still allowing an examination of trends in animal use over time. Lastly, tracking data by species and by purpose of testing (e.g., regulatory, screening, or research and development) will help to identify which NAMs are providing the greatest re-



ductions in animal use and where to prioritize efforts to develop, validate, and increase the use of NAMs.

To continue to improve animal savings and increase human relevance of toxicity tests, new NAM development is needed to cover greater biological space, especially in key study types that use large numbers of animals. NAM approaches to evaluate developmental and reproductive toxicity (DART) endpoints would have a marked impact on animal savings metrics (Rovida and Hartung, 2009; van der Laan et al., 2012). Another opportunity for substantive animal savings is mastering *in vitro* metabolism in NAMs, which would markedly increase their applicability in replacing animal use. In ecotoxicology, effluent toxicity assessments use more fish than chemical hazard assessments (Lillicrap et al., 2016), and, thus, an efficient NAM replacement could generate considerable animal savings. Next generation *in vitro* models (e.g., human stem cell research, organ-on-a-chip) with greater physiological relevance and predictivity will improve confidence in NAM use (Archibald et al., 2015). Furthermore, there is a need for the timely uptake and regulatory acceptance of robust NAMs for human health endpoints after gaining confidence in their relevance to human biology, mechanisms, and domain of applicability, rather than solely relying on their ability to predict the results of animal tests, some of which have shown considerable variability (Luechtefeld et al., 2016; Kleinstreuer et al., 2018; Pham et al., 2020; Clippinger et al., 2021; Rooney et al., 2021).

Greater regulatory acceptance of NAMs should lead to greater animal savings. Toward this goal, laboratories can continue the development of test methods in accordance with the OECD Guidance Document on Good In Vitro Methods Practices (GIVIMP; OECD, 2018b), including documenting test method readiness to further establish confidence in NAMs and understanding scenarios when some *in vivo* follow-up may be required (e.g., Leist et al., 2010; Schmidt et al., 2017). Krebs et al. (2020) proposed several actions that could be adopted to support increased regulatory acceptance of NAMs including depositing standard operating procedures (SOPs) in an accessible public repository, specifying test evaluation methods and pipelines for data processing, and defining uncertainties. However, for reasons from lack of international regulatory acceptance of the animal-free approaches to lack of reviewer awareness of the NAM, regulatory acceptance does not guarantee a decrease in animal use, as was the case, for example, when the EPA's alternate testing framework for classifying the eye irritation potential of antimicrobial cleaning products was found to be underutilized (Clippinger et al., 2016). Thus, a quantitative measure of the implementation of NAMs and animal savings is important to track uptake of NAMs and address any barriers to their use.

From a regulatory perspective, better international harmonization, and mutual acceptance of data (MAD) could lead to greater animal savings (Lillicrap et al., 2016). For example, sometimes a waiver argument is accepted in one country but not in another. Sometimes additional species testing is required in different geographical regions, even when testing in a similar species has been conducted. Agreement on exposure-based waiving and the application of exposure/QIVIVE to judge the relevance of *in vitro* assays could further limit the need for *in vivo* studies. Lastly, defined procedures for the qualification of novel methods, regular training of reviewers of regulatory registration submittals, and

expedited regulatory review of submissions that use NAMs will ensure more rapid evaluation and acceptance of emerging technologies (e.g., EMA Scientific Advice Working Party in Manolis et al., 2011).

In conclusion, this paper proposes one approach to track NAM-based equivalent animal savings based on the number of animals used in *in vivo* studies supplying similar data. Different organizations should tailor this approach for their needs. Some companies have corporate goals around NAM utilization, animal savings or both, and some government agencies have aims or directives to implement alternative methods where available and to track their progress. As we gain additional experience in using NAM data in different scenarios, we will move closer to realizing our shared goal of replacing animal tests with more reliable and relevant NAMs.

References

- Archibald, K., Drake T. and Coleman, R. (2015). Barriers to the uptake of human-based test methods, and how to overcome them. *Altern Lab Anim* 43, 301-308. doi:10.1177/026119291504300504
- Ball, N., Cronin M. T. D., Shen J. et al. (2016). Toward good read-across practice (GRAP) guidance. *ALTEX* 33, 149-166. doi:10.14573/altex.1601251
- Baltazar, M. T., Cable, S., Carmichael, P. L. et al. (2020). A next-generation risk assessment case study for coumarin in cosmetic products. *Toxicol Sci* 176, 236-252. doi:10.1093/toxsci/kfaa048
- Bartlett, D. H. and Silk, S. B. (2016). Office of Laboratory Animal Welfare comments. *Zebrafish* 13, 563-564. doi:10.1089/zeb.2016.1344
- Bhatarai, B., Wilson, D. M., Bartels, M. J. et al. (2015). Acute toxicity prediction in multiple species by leveraging mechanistic ToxCast mitochondrial inhibition data and simulation of oral bioavailability. *Toxicol Sci* 147, 386-396. doi:10.1093/toxsci/kfv135
- Chary, A., Hennen, J., Klein, S. G. et al. (2018). Respiratory sensitization: Toxicological point of view on the available assays. *Arch Toxicol* 92, 803-822. doi:10.1007/s00204-017-2088-5
- Chary, A., Serchi, T., Moschini, E. et al. (2019). An *in vitro* coculture system for the detection of sensitization following aerosol exposure. *ALTEX* 36, 403-418. doi:10.14573/altex.1901241
- Clippinger, A. J., Hill, E., Curren, R. et al. (2016). Bridging the gap between regulatory acceptance and industry use of non-animal methods. *ALTEX* 33, 453-458. doi:10.14573/altex.1601311
- Clippinger, A. J., Raabe, H. A., Allen, D. G. et al. (2021). Human-relevant approaches to assess eye corrosion/irritation potential of agrochemical formulations. *Cutan Ocul Toxicol* 40, 145-167. doi:10.1080/15569527.2021.1910291
- Corvaro, M., Gehen, S., Andrews, K. et al. (2016). GHS additivity formula: A true replacement method for acute systemic toxicity testing of agrochemical formulations. *Regul Toxicol Pharmacol* 82, 99-110. doi:10.1016/j.yrtph.2016.10.007
- Dalvie, D., Obach, R. S., Kang, P. et al. (2009). Assessment of three human *in vitro* systems in the generation of major human excretory and circulating metabolites. *Chem Res Toxicol* 22, 357-368. doi:10.1021/tx8004357

- EC – European Commission (2020). Report from the Commission to the European Parliament and the Council. 2019 report on the statistics on the use of animals for scientific purposes in the Member States of the European Union in 2015-2017. European Commission, Brussels 5.2.2020. Report: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0016&from=EN>; Website: https://ec.europa.eu/environment/chemicals/lab_animals/reports_en.htm (accessed 17.07.2021)
- EFSA – European Food Safety Authority (2014). Modern methodologies and tools for human hazard assessment of chemicals. Scientific report of EFSA. *EFSA J* 12, 3638. doi:10.2903/j.efsa.2014.3638
- EU – European Union (2009). Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on Cosmetic Products. European Union (EU). <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02009R1223-20190813&from=EN#tocId2> (accessed 17.07.2021)
- EU (2010). Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. *OJ L* 276, 33-79. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32010L0063&from=EN> (accessed 17.07.2021)
- EURL ECVAM – EU Reference Laboratory for Alternatives to Animal Testing (2020). EURL ECVAM Recommendation on Non-Animal-Derived Antibodies. doi:10.2760/80554
- Groff, K., Allen, D., Casey, W. et al. (2020). Increasing the use of animal-free recombinant antibodies. *ALTEX* 37, 309-311. doi:10.14573/altex.2001071
- ICCVAM – Interagency Coordinating Committee on the Validation of Alternative Methods (2018). A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States. <https://ntp.niehs.nih.gov/whatwestudy/niceatm/natl-strategy/index.html> (accessed 17.07.2021)
- ICCVAM (2021). ICCVAM Metrics Workgroup. Measuring U.S. Federal Agency Progress Toward Implementation of Alternative Methods in Toxicity Testing. https://ntp.niehs.nih.gov/iccvam/docs/about_docs/iccvam-measuringprogress-feb2021-fd-508.pdf (accessed 17.07.2021)
- 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
- Krebs, A., van Vugt-Lussenburg, B. M. A., Waldmann, T. et al. (2020). The EU-ToxRisk method documentation, data processing and chemical testing pipeline for the regulatory use of new approach methods. *Arch Toxicol* 94, 2435-2461. doi:10.1007/s00204-020-02802-6
- Krieger, S., Auernhammer, T., Hotchkiss, J. et al. (submitted). Mechanistic profiling of inhalation sensory irritants with a computational model for transient receptor potential vanilloid subfamily type 1 (TRPV1).
- Ladics, G. S., Smith, C., Heaps, K. et al. (1995). Possible incorporation of an immunotoxicological functional assay for assessing humoral immunity for hazard identification purposes in rats on standard toxicology study. *Toxicology* 96, 225-238. doi:10.1016/0300-483x(94)02967-y
- Leist, M., Efremova, L. and Karreman, C. (2010). Food for thought considerations and guidelines for basic test method descriptions in toxicology. *ALTEX* 27, 309-317. doi:10.14573/altex.2010.4.309
- Luechtefeld, T., Maertens, A., Russo, D. P. et al. (2016). Analysis of Draize eye irritation testing and its prediction by mining publicly available 2008-2014 REACH data. *ALTEX* 33, 123-134. doi:10.14573/altex.1510053
- Luechtefeld, T., Marsh, D., Rowlands, C. et al. (2018). Machine learning of toxicological big data enables read-across structure activity relationships (RASAR) outperforming animal test reproducibility. *Toxicol Sci* 165, 198-212. doi:10.1093/toxsci/kfy152
- Lillicrap, A., Belanger, S., Burden, N. et al. (2016). Alternative approaches to vertebrate ecotoxicity tests in the 21st century: A review of developments over the last 2 decades and current status. *Environ Toxicol Chem* 35, 2637-2646. doi:10.1002/etc.3603
- Luccio-Camelo, D. C. and Prins, G. S. (2011). Disruption of androgen receptor signaling in males by environmental chemicals. *J Steroid Biochem Mol Biol* 127, 74-82. doi:10.1016/j.jsbmb.2011.04.004
- Manolis, E., Vamvakas, S. and Isaac, M. (2011). New pathway for qualification of novel methodologies in the European medicines agency. *Proteomics Clin Appl* 5, 248-255. doi:10.1002/prca.201000130
- Mansouri, K., Kleinstreuer, N., Abdelaziz, A. M. et al. (2020). COMPARA: Collaborative modeling project for androgen receptor activity. *Environ Health Perspect* 128, 27002. doi:10.1289/EHP5580
- Meigs, L., Smirnova, L., Rovida, C. et al. (2018). Animal testing and its alternatives – The most important omics is economics. *ALTEX* 35, 275-305. doi:10.14573/altex.1807041
- NIH – National Institutes of Health (2021). Guidelines for Preparing USDA Annual Reports and Assigning USDA Reporting Columns. <https://oacu.oir.nih.gov/system/files/media/file/2021-02/a1-usda.pdf> (accessed 23.11.2021)
- 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 – Organisation for Economic Cooperation and Development (2009). Test No. 441: Hershberger Bioassay in Rats: A Short-Term Screening Assay for (Anti)Androgenic Properties. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264076334-en
- OECD (2010). Test No. 429: Skin Sensitization: Local Lymph Node Assay. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264071100-en
- OECD (2015). Adverse Outcome Pathways, Molecular Screening and Toxicogenomics. *OECD Testing of Chemicals*. OECD, Paris, France. <https://www.oecd.org/chemicalsafety/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm> (accessed 17.07.2021)
- OECD (2016). Integrated Approaches to Testing and Assessment (IATA). *OECD Assessment of Chemicals*. OECD, Paris, France. <http://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm> (accessed 17.07.2021)



- OECD (2017). Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to be Used with Integrated Approaches to Testing and Assessment (IATA) for Skin Sensitization. *OECD Series on Testing and Assessment, No. 256*. OECD Publishing, Paris. doi:10.1787/9789264279285-en
- OECD (2018a). Test 443: Extended One-Generation Reproductive Toxicity Study. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264185371-en
- OECD (2018b). Guidance Document on Good In Vitro Methods Practices (GIVIMP). *OECD Series on Testing and Assessment, No. 286*. OECD Publishing, Paris. doi:10.1787/20777876
- OECD (2021). Guideline No. 497: Defined Approaches on Skin Sensitization. *OECD Guideline for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/b92879a4-en
- Pham, L. L., Watford, S. M., Pradeep, P. et al. (2020). Variability in in vivo studies: Defining the upper limit of performance for predictions of systemic effect levels. *Comput Toxicol* 15, 100126. doi:10.1016/j.comtox.2020.100126
- Rooney, J. P., Choksi, N. Y., Ceger, P. et al. (2021). Analysis of variability in the rabbit skin irritation assay. *Regul Toxicol Pharmacol* 122, 104920. doi:10.1016/j.yrtph.2021.104920
- Rovida, C. and Hartung, T. (2009). Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals – A report by the transatlantic think tank for toxicology (t⁴). *ALTEX* 26, 187-208. doi:10.14573/altex.2009.3.187
- Schmidt, B. Z., Lehmann, M., Gutbier, S. et al. (2017). In vitro acute and developmental neurotoxicity screening: An overview of cellular platforms and high-throughput technical possibilities. *Arch Toxicol* 91, 1-33. doi:10.1007/s00204-016-1805-9
- Terry, C., Rasoulpour, R. J., Saghir, S. et al. (2014). Application of a novel integrated toxicity testing strategy incorporating “3R” principles of animal research to evaluate the safety of a new agrochemical sulfoxaflor. *Crit Rev Toxicol* 44, 1-14. doi:10.3109/10408444.2014.910753
- US EPA – US Environmental Protection Agency (2018). Strategic Plan to Promote the Development and Implementation of Alternative Test Methods Within the TSCA Program. June 22, 2018. https://www.epa.gov/sites/production/files/2018-06/documents/epa_alt_strat_plan_6-20-18_clean_final.pdf (accessed 17.07.2021)
- US EPA (2019a). Administrator Wheeler Signs Memo to Reduce Animal Testing, Awards \$4.25 Million to Advance Research on Alternative Methods to Animal Testing. September 10, 2019. <https://www.epa.gov/newsreleases/administrator-wheeler-signs-memo-reduce-animal-testing-awards-425-million-advance> (accessed 26.09.2021)
- US EPA (2019b). Administrator Memo Prioritizing Efforts to Reduce Animal Testing, September 10, 2019. <https://www.epa.gov/research/administrator-memo-prioritizing-efforts-reduce-animal-testing-september-10-2019> (accessed 17.07.2021)
- US EPA (2021). New Approach Methods Work Plan, December 2021. https://www.epa.gov/system/files/documents/2021-11/nams-work-plan_11_15_21_508-tagged.pdf (accessed 17.12.2021)
- US GAO – US Government Accountability Office (2019). Animal Use in Research: Federal Agencies Should Assess and Report on Their Efforts to Develop and Promote Alternatives. GAO-19-629. <https://www.gao.gov/products/gao-19-629> (accessed 17.07.2021)
- van der Laan, J. W., Chapin, R. E., Haenen, B. et al. (2012). Testing strategies for embryo-fetal toxicity of human pharmaceuticals. Animal models vs. in vitro approaches. *Regul Toxicol Pharmacol* 63, 115-123. doi:10.1016/j.yrtph.2012.03.009
- van der Valk, J., Mellor, D., Brands, R. et al. (2004). The humane collection of fetal bovine serum and possibilities for serum-free cell and tissue culture. *Toxicol In Vitro* 18, 1-12. doi:10.1016/j.tiv.2003.08.009
- van der Valk, J., Brunner, D., De Smet, K. et al. (2010). Optimization of chemically defined cell culture media – Replacing fetal bovine serum in mammalian in vitro methods. *Toxicol In Vitro* 24, 1053-1063. doi:10.1016/j.tiv.2010.03.016
- van der Valk, J., Bieback, K., Buta, C. et al. (2018). Fetal Bovine Serum (FBS): Past – Present – Future. *ALTEX* 35, 99-118. doi:10.14573/altex.1705101
- Wilson, D., Wijeyesakere, S. J., Parks, A. K. et al. (2018). Profiling acute oral and inhalation toxicity data using a computational workflow to screen for facile chemical reactivity. *Appl In Vitro Toxicol* 4, 214-219. doi:10.1089/aivt.2017.0041
- Wijeyesakere, S., Wilson, D., Settivari, R. et al. (2018). Development of a profiler for facile chemical reactivity using the open-source Konstanz information miner. *Appl In Vitro Toxicol* 4, 202-213. doi:10.1089/aivt.2017.0040
- Wijeyesakere, S. J., Wilson, D., Auernhammer, T. et al. (2019). Hybrid machine-learning/SMARTS profiling model for mitochondrial inhibition. *Appl In Vitro Toxicol* 5, 196-204. doi:10.1089/aivt.2019.0010
- Wijeyesakere, S. J., Wilson, D. M. and Marty, M. S. (2020). Prediction of cholinergic compounds by machine learning. *Comput Toxicol* 13, 100119. doi:10.1016/j.comtox.2020.100119
- Zhang, F., Bartels, M., Clark, A. et al. (2018). Performance evaluation of the GastroPlus™ software tool for prediction of the toxicokinetic parameters of chemicals. *SAR QSAR Environ Res* 29, 875-893. doi:10.1080/1062936X.2018.1518928

Conflict of interest

The authors declare that they have no conflict of interest to report relative to the contents of this manuscript.

Acknowledgements

The authors would like to thank a number of colleagues for early discussions on this approach and/or their thoughtful review of the tables in this manuscript, including Nick Ball, Joe Chai, Amy Clippinger, Rachel Golden, Matt LeBaron, Troy Marino, Jen Murray, Jessica Scherzer, Raja Settivari, Kristy Swiecicki, and Dan Wilson. The described animal tracking system and this manuscript were funded by Dow as part of its animal welfare program.