Concept Article

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

M. Sue Marty¹, Amanda K. Andrus², and Katherine A. Groff²
Dow, 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 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 human 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 Regulations). The US EPA has proposed a goal to eliminate animal testing by 2035 (US EPA, 2019a, b). In recent years, the Organisation for Economic Cooperation and Development (OECD) has developed internationally accepted test guidelines for a number of NAMs to evaluate dermal irritation, eye irritation/corrosion, and dermal sensitization potential, among
other health effects. International tools, including adverse outcome pathways (OECD, 2015), Integrated Approaches to Testing and Assessment (IATAs; OECD, 2016) and Defined Approaches to Testing and Assessment (DAs; OECD, 2017) describe approaches to apply NAM data with potential application 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 status for regulatory acceptance. In most companies, animal use policies emphasize the 3Rs (i.e., to Replace, Reduce and/or 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 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’s (GAO) highlighted the need to identify metrics to monitor animal use in its recommendation 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 additional 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 in order 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 Figure 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 (AAALAS) Guide. According to the AAALAS 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 include cephalopods, cyclostomes, and fetal mammals in the last third of their development (EU, 2010).

Establishing consistency in tracking animal use

In order 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 animal use internally is not offset by an increase in animal use externally. 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 ecotoxicology 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

Fig. 1: Determining the appropriate number of animals to count towards animal savings depends on how NAM data are used and 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 selection 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 NAM applications 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.

Tab. 1: Definitions of Animals Included in Various Animal Use Tracking Programs

<table>
<thead>
<tr>
<th>Organization</th>
<th>Animal Definition with Inclusions/Exclusions</th>
</tr>
</thead>
</table>
| Dow’s Animal Definition is based on the American Association for Animal Laboratory Science (AALAS) Guide | - “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 or embryos, which is consistent with USDA (NIH, 2015) and Office of Laboratory Animal Welfare (2016)¹,²  
- Invertebrates are not included  
- Animals monitored as part of field studies are excluded |
| US Environmental Protection Agency based on memo on animal use reductions (US EPA, 2019b) | - Focus on mammalian studies (i.e., per 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 | - “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 |

¹The National Institutes of Health Guidelines for Preparing USDA Annual Reports: Viable offspring that are born as a result of an Animal Study Protocol requiring breeding are counted at weaning or when subjected to experimental manipulation if that manipulation occurs earlier than weaning. (USDA, 2009)  
²Office of Laboratory Animal Welfare (OLAW, 2016) guidance³ addresses and recognizes fish as ‘animals’ at “hatching” in accordance with Public Health Service Policy.

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 has a goal to reduce mammalian studies by 30% by 2025 and eliminate all mammalian studies (except by Administrator exemption) by 2035 (US EPA, 2019a, b), and some companies adhere to a goal to conduct tests on animals only to comply with regulatory requirements4.

**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, *in vitro* toxicokinetic metabolism models (e.g., microsomes, S-9 or primary hepatocytes ordered from an external vendor) sometimes 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 or fetal calf serum) 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 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:

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 according to the OECD Guideline 497 “Defined Approaches on Skin Sensitisation” (OECD, 2021), the ‘animal savings’ could be up to 28, 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 5α-reductase inhibitors. 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 5α-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 transactivation 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 a in vivo study.

3. 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 generate 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 level of certainty for decision making (Figure 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 mode-of-action. 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 Cmax, then collect metabolites for identification); thus, it might be reasonable to assign an 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

---

**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.
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:

\[
\text{Total animal equivalents} = \text{No. animals used in in vivo studies} + \text{Equivalent animal savings from NAMs}
\]

\[
\text{% Reduction in animal use} = \left(\frac{\text{Total animal equivalents} - \text{Equivalent animal savings from NAM}}{\text{Total animal equivalents}}\right) \times 100
\]

A third option might be the percentage of toxicity information that comes from NAMs:

\[
\text{% Toxicity data from NAMS} = \left(\frac{\text{Equivalent animal savings from NAM}}{\text{Total animal equivalents}}\right) \times 100
\]

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 by 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 adjusted lower). 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 ability to account for toxicokinetics (absorption, distribution, metabolism and elimination); 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., Bhhatarai 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 absorption, distribution, metabolism or elimination (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.
Tab. 2: Animal Use Reductions due to the Application of *In silico* (Computer-based) NAMs to Predict Toxicity*

<table>
<thead>
<tr>
<th>Endpoint addressed by NAM</th>
<th>Corresponding <em>in vivo</em> test</th>
<th>No. animals <em>in vivo</em></th>
<th>Animal savings using NAM</th>
<th>Rationale for percentage selected</th>
<th>No. animals saved by NAM use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute oral toxicity</td>
<td>OECD 425 (Acute oral toxicity – Up-and-down procedure; options: OECD 420 and 423)</td>
<td>5-8&lt;sup&gt;a&lt;/sup&gt;; 6-12 per Corvaro et al., 2016</td>
<td>30%</td>
<td>• Most potent MOAs identified, but not all relevant targets have been modeled&lt;br&gt;• Predictions are conservative and route agnostic&lt;br&gt;• Evaluation covers parent and predicted highly toxic metabolites&lt;br&gt;• Predicts potential MOA and GHS class</td>
<td>2.1</td>
</tr>
<tr>
<td>Acute dermal toxicity</td>
<td>OECD 402 (acute dermal toxicity – fixed dose procedure)</td>
<td>4-9&lt;sup&gt;b&lt;/sup&gt;; avg = 7; full test 50 per Corvaro et al., 2016</td>
<td>30%</td>
<td>• Similar predictions to acute oral toxicity are used and generally considered conservative</td>
<td>2.1</td>
</tr>
<tr>
<td>Acute inhalation toxicity</td>
<td>OECD 403 (Acute toxic class; other options: OECD 436 and 433)</td>
<td>40-42&lt;sup&gt;c&lt;/sup&gt;</td>
<td>10%</td>
<td>• Not all relevant targets have been modeled&lt;br&gt;• Predictions are conservative and route agnostic&lt;br&gt;• Evaluation covers parent and predicted metabolites&lt;br&gt;• Predicts potential MOA but not GHS class</td>
<td>4</td>
</tr>
<tr>
<td>Acute dermal irritation/corrosion</td>
<td>OECD 404 (Acute dermal irritation/corrosion)</td>
<td>2-3</td>
<td>10%</td>
<td>• Use read-across and mechanistic profiling&lt;br&gt;• Evaluation covers parent&lt;br&gt;• Predict pos or neg, but not GHS class (unless using read-across)&lt;br&gt;• If needed, <em>in vitro</em> follow-up to address uncertainty</td>
<td>0.3</td>
</tr>
<tr>
<td>Eye irritation</td>
<td>OECD 405 (acute eye irritation/corrosion)</td>
<td>1-3</td>
<td>10%</td>
<td>• Use read-across and mechanistic profiling&lt;br&gt;• Evaluation covers parent&lt;br&gt;• Predict pos or neg, but not GHS class (unless using read-across)&lt;br&gt;• If needed, <em>in vitro</em> follow-up to address uncertainty</td>
<td>0.3</td>
</tr>
<tr>
<td>Skin sensitization</td>
<td>OECD 429 or 442A (LLNA) (option: 406)</td>
<td>28&lt;sup&gt;d&lt;/sup&gt;</td>
<td>30%</td>
<td>• Most relevant bioactivities (e.g., facile reactivity) included in models/profilers&lt;br&gt;• Predict pos or neg, but not GHS class (unless using read-across)</td>
<td>8.4</td>
</tr>
<tr>
<td>Filling data gaps for Research Sample Safety Data Sheets (RSSDS) or Material Safety Data Sheets (MSDS)</td>
<td>• OECD 425 (Acute oral)&lt;br&gt;• OECD 404 (Acute dermal irritation/corrosion)&lt;br&gt;• OECD 405 (Eye irritation)&lt;br&gt;• OECD 429 (LLNA)</td>
<td>41&lt;sup&gt;e&lt;/sup&gt;</td>
<td>30%</td>
<td>• Not all relevant targets have been modeled&lt;br&gt;• Screening novel chemical domains&lt;br&gt;• Predictions are conservative and route agnostic; reasonable confidence&lt;br&gt;• Evaluation covers parent and predicted metabolites&lt;br&gt;• Predicts potential MOA but not GHS class for all endpoints (acute oral only)</td>
<td>12.3</td>
</tr>
</tbody>
</table>

| Ecotoxicological QSAR Predictions | | | | | |
| Acute Fish Tox | OECD 203 (fish, acute toxicity test) | 42<sup>f</sup>-49 | 30% | • Models used for PMN and REACH submissions if chemical is in domain<br>• Long-term use with reasonable understanding of reliability | 14.7 |
| Chronic Fish Tox | OECD 210 (fish, early life stage toxicity test) | 560<sup>g</sup> (840)<sup>b</sup> | 30% | | 168 |
| Bioaccumulation (BCF) | OECD 305 (bioaccumulation in fish) | 251<sup>e</sup> | 30% | | 75.3 |

<sup>a</sup> 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™

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

2-3 animals for range-finding study and 2-6 animals in the main study

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.

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

Includes 7 (acute oral) + 3 (dermal irritation) + 3 (eye irritation) + 28 (LLNA) = 41 animals

Definitive test: 7 fish/concentration with minimum of 5 concentrations = 28

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. Does not include fish needed for rangefinder if required (~18-30 fish)

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)

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/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). 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 vivo NAMs to Predict Toxicity*

<table>
<thead>
<tr>
<th>Endpoint Addressed by NAM</th>
<th>Corresponding in vivo Test</th>
<th>No. Animals in vivo</th>
<th>Animal savings using NAM</th>
<th>Rationale for Percentage Selected</th>
<th>No. Animals Saved by NAM use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Endpoints</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3T3 Neutral red uptake (NRU) (OECD* 432) for phototoxicity/cytotoxicity</td>
<td>No formal OECD phototoxicity/cytotoxicity test guideline; Estimate based on other dermal studies</td>
<td>2-3</td>
<td>100%</td>
<td>Equivalent to phototoxicity screen</td>
<td>2</td>
</tr>
<tr>
<td>Reactive Oxygen Species (ROS) for photo-reactivity (OECD 495)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcutaneous electrical resistance test (OECD 430), Membrane Barrier test (OECD 435) or Reconstructed human epidermis (RhE) test (OECD 431) for skin corrosion</td>
<td>OECD 404 (Acute dermal irritation/corrosion)</td>
<td>2-3</td>
<td>100%</td>
<td>Equivalent to skin irritation screen</td>
<td>3</td>
</tr>
<tr>
<td>Reconstructed human epidermis (RhE) test (OECD 439) for in vitro skin irritation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluorescein leakage (OECD 460) or Statens Seruminstitut rabbit cornea (SIRC) cells for eye corrosives or severe irritants (OECD 491)</td>
<td>OECD 405 (acute eye irritation/corrosion)</td>
<td>1-3 (1 if corrosive)</td>
<td>100%</td>
<td>Equivalent to eye irritation screen</td>
<td>3</td>
</tr>
<tr>
<td>Reconstructed human corneal-like epithelium (RhCE) test for eye irritation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bovine corneal opacity test (BCOP: OECD 437), isolated chicken eye test (OECD 438), or Vitrigel (OECD 494) for eye irritation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARE-Nrf2 luciferase KeratinoSens™ or ARE-Nrf2</td>
<td>OECD 429 or 442A (LLNA) (option: 406)</td>
<td>28p</td>
<td>33-50%</td>
<td>2-3 assays are needed to equal in vivo assessment; thus</td>
<td>28</td>
</tr>
<tr>
<td>Assay Description</td>
<td>OECD Code(s)</td>
<td>Duration</td>
<td>Comments</td>
<td>% animal savings</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>Luciferase LuSens test (OECD 442D)</td>
<td></td>
<td></td>
<td>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.</td>
<td>33-50%</td>
<td></td>
</tr>
<tr>
<td>Direct Peptide Reactivity Assay (DPRA) or Amino acid derivative reactivity assay (ADRA) (OECD 442C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>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)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33-50%</td>
</tr>
<tr>
<td><strong>Endocrine Endpoints</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9-18</td>
</tr>
<tr>
<td>ER transactivation assay (ERTA) (OECD 455; binding only in OECD 493)</td>
<td>OECD 440 (uterotrophic bioassay)</td>
<td>18²</td>
<td>50-100%</td>
<td>Uterotrophic can measure agonist and antagonists (test guideline agonists only) ERTA focus on parent cmpd, not metabolites When coupled with ER Expert System (CERAPP), may fulfill requirements to screen for oestrogenicity</td>
<td>9-18</td>
</tr>
<tr>
<td>AR transactivation assay (ARTA) (OECD 458)</td>
<td>OECD 441 (Hershberger assay)</td>
<td>48³</td>
<td>20%</td>
<td>Hershberger can measure AR agonist and antagonists; 5-alpha-reductase inhibitors (5αR I); AR cannot detect 5αR I ARTA focus on parent cmpd, not metabolites</td>
<td>9.6</td>
</tr>
<tr>
<td>Aromatase assay (OPPTS 890.1200)</td>
<td>OPPTS 890.1450 (Female pubertal assay)</td>
<td>60</td>
<td>10%</td>
<td>Female pubertal also can detect steroidogenesis inhibitors, E2, anti-E2, androgens, thyroid; aromatase assay only detects limited subset (i.e., only one enzyme in pathway, although one that is commonly affected by chemicals) Aromatase assessment primarily of parent cmpd</td>
<td>6</td>
</tr>
<tr>
<td>H295R Steroidogenesis assay (OECD 456)</td>
<td>OPPTS 890.1450 or 890.1500 (Female or male pubertal assay)</td>
<td>60</td>
<td>10%</td>
<td>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 cmpd</td>
<td>6</td>
</tr>
<tr>
<td>TPO assay</td>
<td>OPPTS 890.1450 or 890.1500 (Female or male pubertal assay)</td>
<td>60</td>
<td>5%</td>
<td>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 cmpd</td>
<td>3</td>
</tr>
</tbody>
</table>

² Duration varies depending on the assay.
³ Duration varies depending on the assay.

- **ERTA**: Estrogen Receptor Transactivation Assay
- **ARTA**: Androgen Receptor Transactivation Assay
- **OPPTS**: Organization for the Protection of the Human Therapeutic Supply
- **TPO**: Thyroid Peroxidase
Tab. 4: Animal Use Reductions due to *in silico/in vitro* Metabolism or Bioaccumulation

<table>
<thead>
<tr>
<th>Endpoint Addressed by NAM</th>
<th>Corresponding <em>in vivo</em> Test</th>
<th>No. Animals in <em>in vivo</em></th>
<th>Animal savings using NAM</th>
<th>Rationale for Percentage Selected</th>
<th>No. Animals Saved by NAM use</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>in vitro</em> comparative metabolism (IVCM) (mole, rat, rabbit, dog, human)</td>
<td>OECD⁴ 417 Toxicokinetics (Absorption, Distribution, Metabolism, and Elimination; ADME) with inclusion of multiple species for cross-species comparison</td>
<td>Minimum 4 rats, 8 mice and 2 rabbits for cross-species comparison; 2 dogs for AME = 16</td>
<td>50% (max savings = 8 if all species included)</td>
<td>• Probe AME covers 3 species (rat, mouse, rabbit)  • <em>In vivo</em> ADME study also tracks absorption, distribution, time course and elimination of radiotracer  • Only metabolism covered <em>in vitro</em>  • IVCM has 5 species (pools are n=3 individuals/pool)</td>
<td>8 (if all species evaluated)</td>
</tr>
<tr>
<td>Skin Absorption (OECD 428)</td>
<td>Skin absorption (OECD 427)⁴</td>
<td>12</td>
<td>50%</td>
<td>• 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  • <em>In vivo</em> study tracks distribution and elimination of radiotracer</td>
<td>6</td>
</tr>
<tr>
<td><em>in silico</em> GastroPlus modeling (REACH ADME)⁴</td>
<td>OECD 417 Toxicokinetics (Absorption, Distribution, elimination)</td>
<td>8 rats/route x 3 routes = 24 rats; 4 rats/route = 1 animal route; maximum</td>
<td>1 animal route; maximum</td>
<td>• REACH dossier support  • <em>In vivo</em> ADME study also tracks absorption, distribution, time</td>
<td>3 (or 6 if there are gender)</td>
</tr>
</tbody>
</table>
Animal use metrics also should include other aspects of the 3R’s 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) describes 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, 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

### Table 5: Intelligent Study Designs

<table>
<thead>
<tr>
<th>Design Type</th>
<th>Number of Animals</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In silico</strong></td>
<td></td>
<td>GastroPlus™ modeling for IVIVE</td>
</tr>
<tr>
<td><strong>In vitro</strong></td>
<td></td>
<td>fish hepatocytes/S9 for bioaccumulations</td>
</tr>
<tr>
<td><strong>In vivo</strong></td>
<td></td>
<td>OECD 305 Bioaccumulation in fish</td>
</tr>
</tbody>
</table>

- **Metabolism, and Elimination; ADME**
- 12 rats if no gender difference, savings = 3
- course and elimination of radiotracer
  - GastroPlus™ - Route-specific AUC, Cmax, Tmax, Time-course and metabolite predictions; half-life (25% savings without gender evaluation)
- 25% of 4 rats, differences predicted = 1
  - Explore in vivo concentration relative to external dose (one route)
  - Minimum 2 animals/dose with 2 doses to understand in vivo dosimetry

- **OECD Test Guidelines**
  - Located at: [https://www.oecd.org/chemicalsafety/testing/oecgdigestsfortheioecheatings.htm](https://www.oecd.org/chemicalsafety/testing/oecgdigestsfortheioecheatings.htm)
  - Relevant guidelines in Sections 2 Effects on Biotic Systems and Section 4 Health Effects

- **4 animals/dose in a single dose pilot study and two dose levels in the main study (12 animals)** for each route
- **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.
- **Validation of GastroPlus™ reported (Zhang et al., 2018); potential metabolites also identified through TIMES OASIS and/or GastroPlus™**
- **OECD Test Guidelines are located at: [https://www.oecd.org/chemicalsafety/testing/oecgdigestsfortheioecheatings.htm](https://www.oecd.org/chemicalsafety/testing/oecgdigestsfortheioecheatings.htm)**
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 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. 5: Animal Use Reductions due to Intelligent Study Designs

<table>
<thead>
<tr>
<th>Endpoint integrated into another repeat-dose animal study</th>
<th>No. Animals in vivo</th>
<th>Animal savings using Intelligent Designs</th>
<th>Rationale for Percentage Selected</th>
<th>No. Animals Saved by Intelligent Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pig A Gene Mutation Assay (OECD TBD)</td>
<td>30&lt;sup&gt;c&lt;/sup&gt;</td>
<td>100%</td>
<td>Integrated in OECD 407 (28-day), 408 (90-day) or OECD 421/422 (repro screens)</td>
<td>24</td>
</tr>
<tr>
<td>In vivo micronucleus test (OECD 474)</td>
<td>25&lt;sup&gt;c&lt;/sup&gt;</td>
<td>100%</td>
<td>Integrated in OECD 407 (28-day), 408 (90-day) or OECD 421/422 (repro screens)</td>
<td>25</td>
</tr>
<tr>
<td>T-cell dependent antibody response (TDAR)</td>
<td>45&lt;sup&gt;d&lt;/sup&gt;</td>
<td>100%</td>
<td>Integrated in OECD 407 (28-day), 408 (90-day) or OECD 443 (EOGRTS)</td>
<td>45</td>
</tr>
<tr>
<td>Neurotoxicity (OECD 424)</td>
<td>80</td>
<td>100%</td>
<td>Similar to OECD 424</td>
<td>80</td>
</tr>
<tr>
<td>TK/metabolism (blood and urine collection during repeat-dose studies; OECD 417)</td>
<td>8 rats/route x 3 routes = 24 rats; 4 rats/route = 12 rats if no gender difference</td>
<td>50%</td>
<td>Integrated in a 90-d study in Terry et al., 2014</td>
<td>6-12</td>
</tr>
</tbody>
</table>

<sup>a</sup> 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.
<sup>b</sup> n = 6/dose group x 4 groups (3 treated groups + positive and negative control groups); but numbers may differ in the final test guideline
<sup>c</sup> n = 5/dose group x 4 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.
<sup>d</sup> 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

<table>
<thead>
<tr>
<th>Test</th>
<th>Name</th>
<th>Adults</th>
<th>Fetuses&lt;sup&gt;e&lt;/sup&gt;/Pups</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Systemic Toxicity (Oral, Dermal, and Inhalation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OECD&lt;sup&gt;e&lt;/sup&gt; 425, 402, 403</td>
<td>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&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Savings dependent on route(s) for which waiving is applicable</td>
<td>NA</td>
<td>95 or 98 if all routes</td>
</tr>
<tr>
<td>OECD&lt;sup&gt;e&lt;/sup&gt; 407</td>
<td>Read-across using existing hazard information&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OECD&lt;sup&gt;e&lt;/sup&gt; 408</td>
<td>GHS additivity formula for mixtures&lt;sup&gt;d&lt;/sup&gt; (e.g., Corvaro et al., 2016)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OECD&lt;sup&gt;e&lt;/sup&gt; 453</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OECD&lt;sup&gt;e&lt;/sup&gt; 421/422</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeat-Dose Toxicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OECD 407</td>
<td>28-day repeat-dose toxicity study</td>
<td>60&lt;sup&gt;f&lt;/sup&gt;</td>
<td>NA</td>
<td>60</td>
</tr>
<tr>
<td>OECD 408</td>
<td>90-day repeat-dose toxicity study</td>
<td>100&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td>120</td>
</tr>
<tr>
<td>OECD 453</td>
<td>Combined chronic/carcinogenicity study</td>
<td>560&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td>560</td>
</tr>
<tr>
<td>OECD 421/422</td>
<td>Reproductive/Developmental Toxicity Screening study with and without Repeated-dose Toxicity study</td>
<td>80&lt;sup&gt;f&lt;/sup&gt;</td>
<td>520&lt;sup&gt;f&lt;/sup&gt;</td>
<td>600</td>
</tr>
<tr>
<td>OECD 443</td>
<td>Extended One-Generation Reproductive Toxicity Study (EOGRTS)</td>
<td>200&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1040&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1240</td>
</tr>
<tr>
<td>OECD 416</td>
<td>Two-Generation Reproductive Toxicity Study</td>
<td>184&lt;sup&gt;f&lt;/sup&gt;</td>
<td>2080&lt;sup&gt;f&lt;/sup&gt;</td>
<td>2264</td>
</tr>
</tbody>
</table>
**4 Discussion**

This paper describes one approach that can be used to track NAM’s contribution to reducing animal use. The foundation of our metrics is that NAM data have significant value and 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 range 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,
Fig. 3: Our generic template for an Integrated Approach to Testing and Assessment (IATA) of chemicals

As shown, NAMs are important elements to 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 *in vitro* positive 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.

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, 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 in order 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 and thus, aquatic species should be tracked separately. In addition, studies conducted through a consortium may be counted by more than one company and thus, it may be useful to note this for any testing 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 reductions 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 in 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; Kleinsteuere 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 a number of 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


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 Swieciicki, and Dan Wilson. The described animal tracking system and this manuscript were funded by Dow as part of its animal welfare program.