# Summary Minutes

# Scientific Advisory Committee on Alternative Toxicological Methods Meeting

## September 21-22, 2022

## Virtual Meeting

# I. Contents

## II. Location of Background Materials and Presentations

Background materials and presentations for the 2022 Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) meeting are available on the National Toxicology Program (NTP) Past SACATM Meetings page (https://ntp.niehs.nih.gov/events/past/index.html?type=SACATM).

## III. Frequently Used Abbreviations

| | |
|---|---|
| 3Rs | replacement, reduction, and refinement of animal use |
| AFRL | U.S. Airforce Research Laboratory |
| AOP | adverse outcome pathway |
| API | application programming interface |
| CATMoS | Collaborative Acute Toxicity Modeling Suite |
| CATSAC | Chemistry and Acute Toxicology Science Advisory Council (EPA) |
| cHTS | curated high-throughput screening |
| CPSC | U.S. Consumer Product Safety Commission |
| DNT | developmental neurotoxicity |
| DOI | U.S. Department of the Interior |
| EPA | U.S. Environmental Protection Agency |
| EURL ECVAM | European Union Reference Laboratory for Alternatives to Animal Testing |
| FAIR | findability, accessibility, interoperability, and reusability |
| FDA | U.S. Food and Drug Administration |
| FIFRA | Federal Insecticide, Fungicide, and Rodenticide Act |
| GHS | United Nations Globally Harmonized System of Classification and Labelling of Chemicals |
| GIVIMP | Good In Vitro Methods Practices (OECD guidance document) |
| GLP | Good Laboratory Practice |
| HASPOC | Hazard and Science Policy Council (EPA) |
| HSLF | Humane Society Legislative Fund |
| HSUS | Humane Society of the United States |
| IACUC | Institutional Animal Care and Use Committee |
| ICCVAM | Interagency Coordinating Committee on the Validation of Alternative Methods |
| ICE | Integrated Chemical Environment |
| ISTAND | Innovative Science and Technology Approaches for New Drugs (FDA |

| | |
|---|---|
| | program) |
| IVIVE | in vitro to in vivo extrapolation |
| LD50 | dose required to kill half the members of a tested population after a specified test duration or other definition of choice |
| MDDT | Medical Device Development Tools (FDA) |
| NAMs | new approach methodologies |
| NICEATM | NTP Interagency Center for the Evaluation of Alternative Toxicological Methods |
| NIEHS | National Institute of Environmental Health Sciences |
| NTP | National Toxicology Program |
| OECD | Organisation for Economic Co-operation and Development |
| OPERA | Open (Quantitative) Structure–activity/property Relationship App |
| OPP | U.S. Environmental Protection Agency Office of Pesticide Programs |
| PBPK | physiologically based pharmacokinetic |
| PCRM | Physicians Committee for Responsible Medicine |
| PETA | People for the Ethical Treatment of Animals |
| QSAR | quantitative structure–activity relationship |
| SACATM | Scientific Advisory Committee on Alternative Toxicological Methods |
| T-REX | Terrestrial Risk Assessment (EPA resource) |
| TRUST | transparency, responsibility, user focus, sustainability, and technology |
| TSAR | Tracking System on Alternative Methods |

# IV. Attendance

SACATM met virtually on September 21 and 22, 2022. The following individuals participated in the meeting. In addition to participants named below, about 230 people viewed the meeting via webcast on September 21, with about 290 viewing on September 22.

## SACATM Members

Antonio Baines, PhD, North Carolina Central University

Szczepan Baran, VMD, MS, VeriSIM Life

Ellen Berg, PhD, Insitro

Joseph Charest, PhD, Biogen

Amy Clippinger, PhD, PETA Science Consortium International e.V.

K. Nadira De Abrew, PhD, The Procter & Gamble Company (Chair)

Denis Fourches, PhD, Oerth Bio

Sean Gehen, PhD, DABT, Corteva Agriscience

Sue Leary, MS, Alternatives Research and Development Foundation

Adrian Nañez, PhD, Takeda Pharmaceutical Co. Ltd.

Kathryn Page, PhD, DABT, The Clorox Company

Priyanka Sura, DVM, MS, DABT, Gilead Sciences, Inc.

Tamara Tal, PhD, Helmholtz-Centre for Environmental Research UFZ

Misti Ushio, PhD, Moment 3 LLC

## Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Principal Representatives

Brian Berridge, DVM, PhD, DACVP, National Institute of Environmental Health Sciences (NIEHS)

Jessie Carder, MS, U.S. Department of Agriculture

Suzanne Fitzpatrick, PhD, DABT, U.S. Food and Drug Administration Center for Food Safety and Applied Nutrition

John Gordon, PhD, U.S. Consumer Product Safety Commission, ICCVAM Co-chair

Barnett Rattner, PhD, U.S. Department of the Interior

## Other ICCVAM Representatives

Paul Brown, PhD, U.S. Food and Drug Administration Center for Drug Evaluation and Research

Warren Casey, PhD, DABT, NIEHS

William Eckel, PhD, U.S. Environmental Protection Agency Office of Pesticide Programs

Andrew Keebaugh, PhD, U.S. Department of Defense, U.S. Air Force Research Laboratory

Nicole Kleinstreuer, PhD, NIEHS

Monique Perron, ScD, U.S. Environmental Protection Agency Office of Pesticide Programs

Elijah Petersen, PhD, National Institute of Standards and Technology

Shelby Skoog, PhD, U.S. Food and Drug Administration Center for Devices and Radiological Health

## NIEHS Staff

Milene Brownlow, PhD, Designated Federal Official

Robbin Guy

Helena Hogberg, PhD

Mary Wolfe, PhD

Rick Woychik, PhD

### NIEHS Support Contractors

David Allen, PhD (Inotiv, contractor supporting the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods [NICEATM])

John Maruca (Image Associates, contractor supporting the NIEHS Office of Communications and Public Liaison)

Parris Milly (NTT DATA, contractor supporting the NIEHS Office of Communications and Public Liaison)

Nathan Mitchiner (NTT DATA, contractor supporting the NIEHS Office of Communications and Public Liaison)

Steven Morefield, MD (Inotiv, contractor supporting NICEATM)

Catherine Sprankle, MS (Inotiv, contractor supporting NICEATM)

Jonathan Strouse (NTT DATA, contractor supporting the NIEHS Office of Communications and Public Liaison)

### Public

Elizabeth Baker, JD, Physicians Committee for Responsible Medicine

João Barroso, PhD, European Union Research Laboratory for Alternatives to Animal Testing

Ashley Haugen, That Water Bead Lady, Inc.

Joseph Manuppello, MS, Physicians Committee for Responsible Medicine

Sue Marty, PhD, DABT, The Dow Chemical Company

Daniela Ortiz Franyuti, Dr.Sci., F. Hoffmann-La Roche, Roche Innovation Center Basel

Jessica Ponder, PhD, Physicians Committee for Responsible Medicine

Kristie Sullivan, MPH, Physicians Committee for Responsible Medicine

# September 21, 2022

## V. Welcome and Opening Remarks

Dr. Nadira De Abrew, The Procter & Gamble Company, Chair of the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM), called the meeting to order at 10:03 a.m. on September 21. SACATM members and key National Institute of Environmental Health Sciences (NIEHS) staff introduced themselves.

In welcoming remarks, Dr. Rick Woychik, NIEHS and National Toxicology Program (NTP) Director, thanked the SACATM members for their service and noted the importance of the advice they provide. The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) interacts with a wide range of stakeholder groups, so a diversity of expertise is needed on this panel. Dr. Woychik noted the two focus topics on the agenda: defining success toward implementing the ICCVAM Roadmap through metrics relevant to ICCVAM agencies and scientific validation of new

approach methodologies (NAMs). Dr. Woychik expressed his appreciation to the industry stakeholders participating in this discussion and noted the participation of ICCVAM's European Union counterpart organization. He closed by recognizing departing SACATM members and international participants.

ICCVAM Co-chair Dr. John Gordon, U.S. Consumer Product Safety Commission (CPSC), thanked the SACATM members and presenters for their time spent preparing for and participating in the meeting, noting that SACATM members interact with ICCVAM throughout the year. Dr. Nicole Kleinstreuer, NIEHS, stated that SACATM's feedback to ICCVAM and the NTP Interagency Center for the Evaluation of Alternative Methods (NICEATM) guides their activities through the year. Dr. Warren Casey, NIEHS, noted the evolution of the SACATM committee over the last 10 years and the role of NICEATM and ICCVAM in making this a productive meeting.

Dr. Milene Brownlow, NIEHS, the SACATM Designated Federal Official, read the conflict-of-interest statement and reviewed meeting logistics.

## VI. Major ICCVAM Accomplishments in 2022

Dr. Gordon provided an overview of ICCVAM activities over the last year to advance the 3Rs: replacement, reduction, and refinement of animal use in testing. He reviewed the charges, activities, and publications of each active ICCVAM workgroup.

- The **Acute Toxicity Workgroup** has published papers over the past four years describing U.S. and international acute toxicity testing needs, in vitro and in silico approaches for predicting acute toxicity, and an evaluation of an additivity approach for predicting acute systemic toxicity of mixtures. Current efforts focus on developing models for acute inhalation toxicity.
- The **Consideration of Alternative Methods Workgroup** is working with stakeholders to develop a white paper on fostering and considering the use of NAMs to reduce animal use. A key element of this is the role funding opportunities play in promoting and communicating availability of NAMs. The activities of this workgroup will be discussed in more detail in Session II.
- The **Ecotoxicology Workgroup** published a survey of U.S. agency ecotoxicity information needs and uses. This information provides the background needed for identifying relevant NAMs, including identification of tests for potential replacement and the policy and regulatory context in which data from those tests are used. The workgroup is currently reviewing available NAMs for acute fish toxicity.
- The **In Vitro to In Vivo Extrapolation (IVIVE) Workgroup** published a review of IVIVE methods and models used by member agencies. The manuscript presents case studies of how IVIVE has been used in risk assessment. The workgroup is currently interacting with the Organisation for Economic Co-operation and Development (OECD) to advance international harmonization of the use of IVIVE.
- The **Nanomaterials Workgroup** published a review of U.S. federal agency and international regulatory information requirements and testing needs. The paper also describes the extent to which alternatives to animal testing can be used to

fulfill these needs. Having completed its charge, the workgroup has transitioned to an expert group. It is not actively meeting but members are still able to communicate and share information.

- The **Validation Workgroup** was established to update the 1997 ICCVAM "Validation and Regulatory Acceptance of Toxicological Test Methods[1]." The update will consider well-established guidance documents that have been published since then, as well as advances in technologies and best practices. This group's activities will be discussed in more detail in Session III.

Dr. Gordon summarized 2022 ICCVAM public interactions.

- Approximately 350 attendees from 22 countries viewed the annual Communities of Practice webinar, which focused on non-animal approaches for neurotoxicity, including developmental neurotoxicity (DNT)[2] (January 2022).
- ICCVAM and NICEATM participated in many activities at the Society of Toxicology meeting[3] (March 2022). Presentations included a continuing education course, a satellite meeting, six platform sessions, and 18 poster presentations.
- Over 100 attendees viewed the ICCVAM Public Forum, which featured member agencies' presentations about their activities during the year and enabled interactions with the public[4] (May 2022).

***Clarifying questions and comments:*** There were no clarifying questions.

## Public Comments

A written public comment was submitted for this section from the Humane Society of the United States (HSUS)/Humane Society Legislative Fund (HSLF).[5]

### *Oral Public Comments*

Ms. Elizabeth Baker, representing the Physicians Committee for Responsible Medicine (PCRM), commended NICEATM and ICCVAM for progress and leadership, and noted the high level of ICCVAM publication and communication activities. PCRM is concerned that Dr. Kleinstreuer remains in an acting role as NICEATM Director and feels that this situation undermines the stability of ICCVAM. Ms. Baker recognized the importance of the Ecotoxicity Workgroup's publication. Availability of this document clarifies the use of data for decision-making and makes suggestions on avoiding animal tests that are not used for decision-making. She encouraged engagement in other similar projects. Agencies should consistently review testing needs and identify areas where decisions can be made without testing. Ms. Baker recognized the accomplishments of the U.S. Environmental Protection Agency (EPA) and the U.S. Food and Drug Administration (FDA). PCRM is pleased to see financial support for advancement of NAMs within FDA

[1] Available at https://ntp.niehs.nih.gov/iccvam/docs/about_docs/validate.pdf.
[2] Available at https://ntp.niehs.nih.gov/go/commprac-2022.
[3] Available at https://ntp.niehs.nih.gov/go/niceatm-sot22.
[4] Available at https://ntp.niehs.nih.gov/go/iccvamforum-2022.
[5] Written public comments are available at https://ntp.niehs.nih.gov/events/past/index.html?type=SACATM (click the link "Meeting Materials" in the far-right table column).

in the current federal budget. However, PCRM is concerned about provisions for parallel animal testing for validation of NAMs, especially in the event that results do not align, and would prefer to see human data being used for comparison with NAMs when possible. PCRM also welcomes FDA's recent announcement that a letter of intent has been accepted for a tool to be qualified through the Innovative Science and Technology Approaches for New Drugs (ISTAND) program. Ms. Baker closed by thanking the organizers for the opportunity to participate in this meeting and noted that PCRM would also welcome the opportunity to participate in ICCVAM workgroup meetings.

### *Comments from Designated SACATM Discussants*

Discussants for "Major ICCVAM Accomplishments in 2022" were asked to consider the following questions:

- What other test methods or endpoints would you consider to be best suited for prioritization by ICCVAM?
- What suggestions do you have for potential topics for a future ICCVAM Communities of Practice webinar?

Dr. Amy Clippinger, PETA Science Consortium International e.V., first discussant, noted the advances made in the last year on evaluation of alternatives for acute toxicity endpoints. She encouraged acknowledgement of the possibility that unrealistic expectations are being set for NAM reproducibility due to the variability of the animal data to which NAMs are being compared. She reiterated the concern expressed by Ms. Baker about the FDA proposal for conducting new animal tests for NAMs validation. There may also be value in identifying and merging NAM classification categories with the same practical effect (i.e., categories that result in the same personal protective recommendations).

Regarding which test methods or endpoints are best suited for prioritization by ICCVAM, Dr. Clippinger encouraged support of the OECD project to develop integrated approaches for testing and assessment for fish toxicity. ICCVAM engagement with this effort should increase its acceptance by U.S. regulators. ICCVAM agencies should also accept data from the in vitro OECD guideline test for acute fish toxicity[6], and share information with stakeholders about any additional data needed to satisfy regulatory requirements. She encouraged agencies to look for opportunities to waive animal tests or replace them with more informative alternatives. Dr. Clippinger recommended that agencies continue to examine fundamental aspects of in vitro testing including establishing processes for high-quality measurements, advancing use of non-animal components such as alternatives for fetal bovine serum and antibodies, and promoting clear communication about protocols and reagents so studies can be easily replicated. Agencies should have webpages devoted to lists of accepted NAMs.

Dr. Clippinger proposed the following ideas for future ICCVAM Communities of Practice webinars:

- Agency presentations about informational needs or the processes they use for

[6] Test No. 249: Fish Cell Line Acute Toxicity – The RTgill W1 Cell Line Assay; available at https://www.oecd-ilibrary.org/environment/test-no-249-fish-cell-line-acute-toxicity-the-rtgill-w1-cell-line-assay_c66d5190-en.

risk assessment.

- Case studies that improve understanding of use and interpretations of new methods.
- Panel discussions on relevant topics that can generate new ideas.
- Funding resources.
- International harmonization.

Referencing NICEATM's current webinar series on population variability, Dr. Clippinger recommended that ICCVAM and NICEATM broaden their reach to a wider audience. She specifically noted the need to reach academics, diverse nongovernmental organizations, and small companies that might not be connected to trade associations. She closed by encouraging the attendees to share what they learn with colleagues and encourage participation and engagement with future ICCVAM events.

Dr. Misti Ushio, Moment 3 LLC, second discussant, noted that Dr. Gordon's review of the publications issued by the ICCVAM workgroups highlights the potential impact of NAMs across multiple agencies and regulatory applications. While welcoming the FDA's announcement of funding for validation on NAMs, she agreed with points that had been made about comparing NAMs to variable animal data and the need for finding more human-relevant alternative approaches to validation. She also encouraged consideration of the barriers to acceptance of new methods, and whether they might be addressed by education about interpretation of NAMs results and increasing familiarity with new types of data.

### *Additional SACATM Comments*

In response to Dr. Clippinger's comments about animal variability, Dr. Gordon agreed that animal tests and NAMs need to be held to the same standard, especially considering the variability of the animal data. Ongoing work aims to identify sources of uncertainty in NAMs and to define standards for quality systems. He acknowledged the importance of outreach, especially to engage ICCVAM agencies' regulatory affairs departments, a point that Dr. De Abrew concurred with. Finally, he agreed with Ms. Baker that Dr. Kleinstreuer needs to be made the permanent director of NICEATM.

Dr. Kleinstreuer appreciated the suggestions for future focus areas and noted that variability of reference data would be a major focus later in this meeting.

Dr. Joseph Charest, Biogen, expressed interest in identifying evidence-based data sets that would provide a basis for demonstrating relevance of NAMs to human biology. Neurotoxicity represents an area where NAMs can improve on existing models that don't work well. He suggested that ICCVAM consider identifying approaches that could be used for both toxicity and efficacy testing of methods. The developer audience needs information about validation and qualification of new methods; the FDA ISTAND program is an example. The earlier developers know what the context of use for a new system will be, the better chance they have of successfully developing a method to meet that context of use. Finally, he agreed with previous comments about the need for ICCVAM and NICEATM to broaden their audience.

Dr. Kathryn Page, The Clorox Company, commended the ICCVAM workgroups on the

number and quality of publications they have issued and encouraged them to continue. However, these publications don't replace clear communication from agencies about methods they will accept. She encouraged prioritization of efforts to eliminate tests that use the most animals. She agreed with Dr. Clippinger's comment about broadening education and outreach efforts and suggested a future Communities of Practice webinar on how agencies will accept NAM data.

Dr. Tamara Tal, Helmholtz-Centre for Environmental Research UFZ, commented that prioritization of activities should consider public perceptions of, and expectations for, chemical safety. Specifically, expectations for chemical safety with regards to DNT are not being met because so few chemicals have been thoroughly tested for this endpoint. The human-based in vitro testing batteries currently under consideration by OECD could improve on current animal models. Methods that could fill this gap should be prioritized at least as much as replacements for tests that use many animals. Assessing NAMs use for complex endpoints such as these would be a good topic for a future Community of Practice webinar, as well as anchoring biological relevance.

Dr. Ellen Berg, Insitro, agreed that the DNT endpoint should be prioritized, as well as cardiovascular toxicity due to the abundance of high-quality clinical data. A good topic for a future Communities of Practice webinar might be a program on data analysis aimed at a broad audience; current webinars on these topics seem to be aimed at experts rather than at people new to these methods.

# VII. Implementing the Strategic Roadmap: Incorporation of Alternatives and Associated Metrics

## Introduction

In his introduction to this session, Dr. Casey noted that, while metrics is a perennial topic of interest, only now can a meaningful discussion of the topic take place because of the existence of validated methods to replace animal use. The principles put forth in the 2018 ICCVAM Strategic Roadmap[7] are useful for guiding discussions of metrics. One of the Strategic Roadmap's key concepts is connecting test method developers with end-users early in the development of new methods. More flexible approaches to validation have enabled greater progress in the acceptance of NAMs, to the point where we can now look at how to measure progress.

A second Roadmap key concept encourages adoption of new methods by agencies and involves three main activities: providing clear language about the acceptance of NAMs, international harmonization, and identifying appropriate metrics, the focus of today's discussion.

Acceptance of NAMs requires a fit-for-purpose validation approach that reflects the needs and criteria of the individual offices within regulatory agencies. This is also true about metrics, which need to consider agencies' differing regulatory statutes and the differing ways that stakeholders report to their regulators. Successful measurement of progress toward acceptance of NAMs requires cooperation and interaction between agencies and their regulated industries. The following talks presented case studies from

[7] Available at https://ntp.niehs.nih.gov/go/natl-strategy.

both agencies and stakeholders relevant to measuring animal use and implementation of NAMs. Dr. Casey closed by noting the availability of the ICCVAM 2020-2021 Biennial Progress Report and demonstrated how users can filter on articles specific to metrics.[8]

***Clarifying questions and comments:*** There were no clarifying questions or comments.

## Communicating Progress in Advancing Alternative Methods for Regulatory Use at the FDA

Dr. Paul Brown, FDA Center for Drug Evaluation and Research, began his talk by reviewing the breadth of FDA's activities. The various FDA centers operate under different procedures and requirements for which specific data and information are needed. FDA's commitment to implementing alternatives to animal testing goes back to their statement in 1988 that discontinued the requirement of LD50. Other activities include reduced animal use through international harmonization, participation in ICCVAM and OECD workgroups that have developed and evaluated alternative methods, and development of methods through collaborations with other federal agencies and industry. In 2017, FDA issued its Predictive Toxicology Roadmap[9], and now has a website devoted to alternatives[10].

Dr. Brown noted that, to advance the 3Rs, FDA is making information available about research and policy activities that impact the 3Rs and provides and tracks training in this area. Communication channels include the ICCVAM Biennial Report and annual reports on progress implementing the Predictive Toxicology Roadmap. The "Advancing Alternative Methods at FDA" website, which is managed by the FDA Alternative Methods Working Group, provides links to FDA publications, guidance, and other resources relevant to alternative methods. The Working Group has a webinar series that allows developers to make presentations about their methods directly to FDA staff; developers can also connect with FDA through the website. Webinars are currently scheduled through the end of 2022; the 15 webinars over the past year have included presentations on organs-on-chips and tissue models for developmental toxicity, neurotoxicity, pharmacokinetics, and veterinary applications.

FDA also advances alternatives to animal testing through tool development programs. The Center for Devices and Radiological Health publishes a list of qualified tools, funding opportunities, and other resources on its Medical Device Development Tools webpage[11]. The Center for Drug Evaluation and Research's ISTAND program[12] is designed to expand the availability of drug development tools to include new technologies and will produce a list of qualified methods to be published on the website. Expected funding will allow expansion of the qualification of alternative methods, including guidance to stakeholders and applied research. Potential guidance would

[8] Available at https://ntp.niehs.nih.gov/iccvamreport/2021/tags/?topic=Metrics.
[9] Available at https://www.fda.gov/science-research/about-science-research-fda/fdas-predictive-toxicology-roadmap.
[10] Available at https://www.fda.gov/science-research/about-science-research-fda/advancing-alternative-methods-fda.
[11] Available at https://www.fda.gov/medical-devices/medical-device-development-tools-mddt.
[12] Available at https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/innovative-science-and-technology-approaches-new-drugs-istand-pilot-program.

address the qualification process and issues specific to safety or developmental areas or provide guidance for considerations of microphysiological systems. FDA also hopes to make its Alternative Methods website more user-friendly by providing more context about the information found there, collaborations with external groups, and links for regulated industry. FDA would also like to produce more publications relevant to alternative methods.

***Clarifying questions and comments:*** Dr. Page asked if it would be possible for FDA to track animal use in applications for which alternatives are available. Dr. Brown replied that there aren't that many animal tests for which FDA requires data that have accepted alternatives. One application where there is an accepted alternative is ocular irritation, but there's no system in place to track that sort of data. Dr. Adrian Nañez, Takeda Pharmaceutical Co., asked if there was a way that FDA could work with industry to improve availability of information about experiences with alternative methods. Dr. Brown responded that FDA interacts with a couple of organizations on a regular basis. Some of these activities bring in diverse stakeholders resulting in a good exchange of information. FDA also interacts internationally, which is critical for reducing animal use globally. He noted the importance of publishing the outcomes of these interactions.

## CPSC Metrics on New Approach Methods Synopsis

Dr. Gordon reviewed the diversity of activities performed by CPSC and indicated that chemical safety evaluations conducted under the Federal Hazardous Substances Act do not require use of specific tests. The 2012 CPSC animal testing policy encourages manufacturers to use alternatives to animal testing in required safety assessments. CPSC issued a guidance document in April 2022, currently available on Regulations.gov[13]; the CPSC website will be updated this fall to incorporate this information. Emphasizing that CPSC does not do or require testing, Dr. Gordon noted that CPSC will be looking at the number of instances where non-animal data have been used for labeling determinations, as well as the number of methods that have been evaluated by CPSC for this testing.

***Clarifying questions and comments:*** Dr. Antonio Baines, North Carolina Central University, asked whether CPSC does anything to ensure that stakeholders are making their best efforts to reduce animal use in testing. Dr. Gordon reiterated that CPSC doesn't require any specific tests be done to meet its requirements and noted that they work directly with regulated stakeholders to clarify data needs. The new guidance document includes a nomination form for new methods that specifies data requirements, which will be evaluated on a case-by-case basis.

## Animal Reduction Metrics Used by EPA OPP

Dr. Monique Perron, EPA, began her talk by describing the scope of the EPA Office of Pesticide Programs (OPP), whose activities are guided by the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA). Studies required under FIFRA are specified in the Code of Federal Regulations and reflect the context in which the product will be used, for example, for food products vs. non-food products. Registration of a new conventional pesticide requires a substantial amount of testing that can use over 10,000

[13] Available at https://www.regulations.gov/document/CPSC-2021-0006-0010.

animals. However, OPP is working to advance the use of NAMs in regulatory risk applications. FIFRA provides flexibility to consider waivers and alternatives, and these decisions are guided by the 2016 Guiding Principles for Data Requirements[14]. Two OPP committees provide advice on waiver requests:

- The Chemistry and Acute Toxicology Science Advisory Council (CATSAC) focuses primarily on acute studies, and its decisions are informed by available EPA[15] and OECD[16] guidance.
- The Hazard and Science Policy Council (HASPOC) considers longer-term studies under EPA guidance[17]. The overall goal is to identify what data are needed to make a regulatory decision to avoid unnecessary testing and expense.

In 2020 EPA issued its NAMs work plan[18], which outlined objectives and strategies for animal reduction; one of the provisions in the work plan is measuring progress. While waivers are tracked and published in annual reports required by the Pesticide Registration Improvement Act, a newer webpage[19] provides information on reducing animal use with a subpage that focuses on metrics[20]. Data on animal reductions since 2018 are summarized there, as well as estimates of cost savings. Some of the data reflects animal savings from acute dermal testing waivers and implementation of in vitro metrics. In summary, Dr. Perron noted that data are reported every year, progress can be tracked by comparison with historical values, specific statistics are associated with specific guidance documents, and these activities reflect extensive collaboration.

***Clarifying questions and comments:*** Dr. De Abrew asked about how mixtures are considered in these evaluations. Dr. Perron replied that HASPOC mostly focuses on individual chemical data, but CATSAC evaluations include a lot of mixtures. EPA's guidance on waivers for acute dermal testing was first implemented for formulations. Dr. Page observed that time saved on review of animal data might be worth tracking and wondered if EPA might be able to establish standard timings for this purpose. She also asked if it would be possible to gather data from registrants about animals used in preliminary studies as part of a submission package. In response to the first question, Dr. Perron noted that the Craig et al. paper that summarized the impact of waivers granted by HASPOC[21] estimated the cost of a contractor review of a submission but didn't consider EPA time spent, which would be challenging to determine. Time saved for reviews would also vary depending on the type of study and the effect observed. Regarding animal use in preliminary studies, one approach would be to ask registrants for that information. It might also be possible to make some assumptions about animal use for preliminary studies for a specific kind of registration and build estimates from

[14] Available at https://www.epa.gov/pesticide-registration/guiding-principles-data-requirements.
[15] Available at https://www.epa.gov/pesticide-registration/bridging-or-waiving-data-requirements.
[16] OECD Series on Testing and Assessment No. 237, available at https://www.oecd.org/env/ehs/testing/mono%202016%2032.pdf.
[17] Available at https://www.epa.gov/pesticide-registration/determining-toxicology-data-requirements.
[18] Available at https://www.epa.gov/chemical-research/new-approach-methods-work-plan.
[19] Available at https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/strategic-vision-adopting-new-approach.
[20] Available at https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/strategic-vision-adopting-new-approach-0.
[21] Available at https://doi.org/10.1016/j.yrtph.2019.104481.

that. Dr. Sean Gehen, Corteva Agriscience, asked about how EPA links a NAM submission to metrics on animal savings. Dr. Perron replied that when EPA uses NAM data to support a study waiver, it will be documented within the Agency memos and described two cases where in vitro data were used to support waiver applications.

## DOI: Potential Metrics to Track and Encourage Use of Alternative Methods in Ecological Research and Testing

Dr. Barnett Rattner, U.S. Department of the Interior (DOI), described the mission of the DOI, which includes protecting ecosystems, natural resources for recreation (e.g., viewing, photography) and consumptive use (e.g., fishing, hunting). DOI activities include research, biomonitoring, damage assessment, diagnostics, and approval of specific types of products. A U.S. Government Accountability Office report on animal use in federal research suggested that DOI could improve its accounting of animals used relative to these activities. Challenges inherent to this sort of accounting include the fact that birds, cold-blooded vertebrates, and invertebrates, which are all important to DOI activities, are not subject to Animal Welfare Act reporting requirements. To address the concern raised by the Government Accountability Office report, DOI has been identifying activities that can be considered "toxicity testing" and who has responsibility for these activities. The overall goal is to reduce animal use in ecotoxicological research and testing without compromising data needed for decision-making related to conservation and management of natural resources. A survey of DOI toxicity testing and animal use alternatives revealed that most activities can be characterized as biomonitoring, with a little testing, under a variety of statutes. Alternatives used include in vitro and in silico approaches, but challenges to applying these include the diversity of organisms that DOI needs to test and monitor and the unique regulatory contexts that these activities are conducted in, such as providing data to the Department of Justice to support natural resource damage claims. One idea for DOI metrics involves tracking training; specifically, both the number of presentations given and the number of scientists involved. Another metric being considered is gathering empirical data related to the use of animals and alternatives in ecotoxicological research and testing, including data from Institutional Animal Care and Use Committee (IACUC) reporting, publications, and data releases. They are also considering initiating a specific reporting requirement for ecological research and toxicity testing.

***Clarifying questions and comments:*** Dr. Tal asked if there are programs in place for noninvasive biomonitoring, especially for larger animals. Dr. Rattner replied that DOI has some, but their use is limited and noted that blood draws are used rather than euthanization whenever possible. Ms. Sue Leary, Alternatives Research and Development Foundation, asked for clarification of the Department of Justice data requirements. Dr. Rattner explained that these occur in the context of natural resource damage assessments, which can include biochemical, pathological, or population effects. Some of these data are used as evidence in court in environmental cases, including the Exxon Valdez and Deepwater Horizon cases.

## Industry Case Study: Tracking NAMs Impacts on Animal Use

Dr. Sue Marty summarized an effort within The Dow Chemical Company to track NAM

impacts on animal use, described in detail in Marty et al. 2022[22]. In this context, she emphasized two key points:

- Dow believes that all NAMs provide useful information for internal decision-making and thus have an impact on animal savings.
- This is an initial approach that will certainly be improved upon over time.

Dow considers NAMs to include any non-animal approaches for testing and assessment, including computer models, read-across, and in vitro or in chemico laboratory methods. Dr. Marty also defined what Dow considers an animal and the concept of equivalent animal savings. The goals of the analysis are to monitor uptake of NAMs use over time, to justify resources spent on NAMs development, and to identify areas where NAMs development is still needed. The first step needed in an effective animal use tracking program is to establish a baseline; for Dow, this includes tracking both in-house animal use and use of animals by partners such as contract research organizations and consortia. They also track study type, separate mammalian and non-mammalian animal use, and compile data on how study requirements vary from year to year. Establishing a multi-year average for animal use can avoid distortions caused by outliers, such as a year in which an unusual number of reproductive studies were done, each of which would use a large number of animals. Estimating animal savings from NAMs depends on how data are used and the level of uncertainty of those data. A NAM's impact on animal savings might be lower in an early stage of development, where decisions on candidate chemistries and hazard profiles are being made. On the other hand, a NAM that can be used in a later stage of development to replace or justify a waiver for a regulatory animal test can have a large impact on animal use. However, even without regulatory acceptance, NAM data have value.

Dr. Marty then presented several tables from her paper, describing animal savings from in silico and in vitro safety and toxicity assessments and study waiving. They also tracked animal savings achieved through "intelligent design:" the design of studies to collect additional endpoint data to avoid having to run an anticipated future study. Individual data in these tables included endpoint, corresponding in vivo test, and animal savings. A decision tree Dow uses for animal savings considers the availability of a NAM for the endpoint of interest and the extent to which it replaces the animal study for regulatory purposes. Most frequently the NAM partially replaces the animal study, and Dr. Marty explained how animal use can be calculated in those cases. In conclusion, Dr. Marty stressed that Dow feels that NAMs have value for internal decision-making, and that value will grow as more NAMs become available.

***Clarifying questions and comments:*** Dr. Priyanka Sura, Gilead Sciences, asked Dr. Marty to clarify how Dow establishes a consistent baseline of animal use, with regards to what is included and excluded. Dr. Marty replied that while Dow has been tracking animal use internally for a long time, tracking animal use in studies done by consortia is more challenging. She also emphasized the importance of lab groups working with different types of NAMs, for example cheminformatics and in vitro NAMs, to share information on the context of their studies so that NAMs that are essentially informing on the same endpoint are not counted twice in terms of animal savings. In response to Dr.

[22] Available at https://doi.org/10.14573/altex.2107211.

Nañez's question on insights gained on the impact of NAMs because of this study, Dr. Marty commented that interactions between research and development teams and toxicologists had been strengthened earlier in the drug development process so that compounds that advance have a better safety profile and require fewer tests. The data are also being used to predict aspects of test design and collect more information within studies.

## Industry Case Study: A Data-driven Decision-making Framework for the Selection, Application, and Development of Advanced In Vitro Models for Preclinical Drug Development

Dr. Daniela Ortiz Franyuti described how Roche is working toward exchanging animal models for alternatives. The drug industry has a great need to make better predictions. A limitation of the current system is the physiological difference between humans and animals. The biological diversity of humans is also not well represented in animal studies. Other issues include complex modes of action of new therapeutics and the poor reproducibility of clinical studies. She summarized the characteristics of studies that are more likely to be replicated, which include good reporting standards, management of data and metadata, and documentation of analysis. Unfortunately, key decisions need to be made before clinical trials. She described a case study in which Roche found an opportunity to put a molecule into clinical trials without animal studies needed, due to differences between human and animal major histocompatibility complex molecules. Instead, Roche did a set of organ-specific in vitro tests, which generated a large amount of data that needed to be interpreted and preserved for the future. This calls for good scientific data management, which requires application of FAIR data management principles: findability, accessibility, interoperability, and reusability. Such data management practices will support a framework where research data supports clinical data as well as the reverse.

***Clarifying questions and comments:*** Dr. Berg noted that the value of in vitro data is improved by a large reference database across drugs and asked how Roche is working toward data sharing. Dr. Ortiz Franyuti said Roche would like to publish their data and are working on doing that in a way that will also protect patient privacy and intellectual property. Dr. Sura asked how Roche applies in vitro tests in a way that accounts for compensatory mechanisms in the whole body. Dr. Ortiz Franyuti agreed that this is a complex problem, especially for immunocompetent models. The technology to do this is maturing but there are still limitations. Currently Roche is just looking at organ effects individually or in limited combinations.

## Consideration of Alternative Methods Workgroup

Ms. Jessie Carder, U.S. Department of Agriculture, summarized the background and goals of the ICCVAM Consideration of Alternative Methods Workgroup. Despite requirements, there is little incentive for investigators to replace animal use with NAMs. This workgroup is considering activities needed to encourage this shift, which might include identifying effective incentives or opportunities for modification of data requirements. While the workgroup's primary focus is on activities related to toxicology testing, their findings could be applied more broadly to research activities. Ms. Carder reviewed the specific charges of the workgroup, including production of a white paper on

approaches to the use of NAMs, collaboration with international counterparts to share ideas and promote harmonization, highlighting grant opportunities to advance development of alternatives, improving communication to promote use of NAMs, and encouraging ICCVAM agencies to promote avenues where NAMs can be better considered and leveraged. The workgroup's focus in 2022 has been on convening small group discussions with stakeholders that have provided a diversity of viewpoints from different industries. She reviewed stakeholder discussion questions, which addressed topics such as barriers to implementing NAMs, appropriate use of NAMs, success stories, and suggestions of communications activities that would promote use of NAMs. These discussions will continue through the end of 2022, with the goal of compiling learnings into a publication to be issued in 2023.

***Clarifying questions and comments:*** In response to Dr. Berg's question about barriers, Ms. Carder shared that lack of consistency of regulatory requirements was a concern of the agrochemical companies, as well as cost of developing NAMs and lack of industry participation.

## Public Comments

Two written public comments were submitted for this section, on behalf of HSUS/HSLF and That Water Bead Lady, Inc.

### *Oral Public Comments*

Ms. Ashley Haugen, representing That Water Bead Lady, Inc., cautioned against dependence on non-human models at the expense of considering real-world human data, which are needed to prevent a medical finding to be prematurely considered as accepted science. Making decisions based on findings limited to specific organ types or specific models risks overlooking "unknown-unknowns." These may include non-target organ effects, especially when there is long-term exposure. Current poison control data are inconsistent and obtained through voluntary reporting, and thus do not provide a good basis for risk assessment. Better real-world human data reporting systems are needed to support validation of NAMs.

Mr. Joseph Manuppello, representing PCRM, presented an example of how counting animals was found to be an effective metric. The goal of these studies[23] was to compare the numbers of animals used to the minimum recommended by the relevant guidance and thereby identify opportunities to reduce animal use within the existing framework of guidance. PCRM counted animals used in carcinogenicity studies summarized in five years' worth of FDA New Drug Applications. In 109 carcinogenicity studies, over 65,000 animals were used, which greatly exceeded the number suggested by European Medicines Agency guidance. Some of the excess animal use appears to be due to the use of dual control groups, which could be eliminated if FDA requirements were harmonized with European Medicines Agency practice. In addition, in toxicokinetic studies researchers tended to add more mice to treatment groups to compensate for expected mortality during blood draws. This could be reduced by the implementation of microsampling, which has a refinement benefit in being less stressful for animals

[23] Reported in Manuppello et al. 2020, https://doi.org/10.1016/j.yrtph.2020.104666, and an additional manuscript by Manuppello et al. in preparation.

as well.

***Clarifying questions and comments:*** Dr. Nañez asked Mr. Manuppello whether, when referring to microsampling, he was specifically referring to the dried blood-spot technique. Mr. Manuppello said generally yes, although he is aware of new methods that avoid the problems of the dried blood-spot technique.

### *Comments from Designated SACATM Discussants: Implementing the Strategic Roadmap: Incorporation of Alternatives and Associated Metrics*

For this session, discussion questions were broken into subtopics and assigned to specific discussants. Discussants for the subtopic of "Metrics Case Studies (Agencies)" were asked to consider the following questions:

- How can ICCVAM or NICEATM encourage groups or entities to share information on metrics reporting?
- What suggestions do you have for other types of information that could be collected to indicate the impact of NAMs?

Ms. Leary, first discussant, noted that this question might be simplified with the implementation of requirements to record animal use like those in place in other countries. She recognized that such an effort is outside the scope of this group but encouraged participants in this meeting to consider providing such information voluntarily, and regulators to encourage and facilitate their regulated communities to do so. NAMs uptake and NAMs use is key to making progress in this area. Ms. Leary felt that the examples Dr. Marty presented on using metrics would be useful for attendees.

Dr. Baines, second discussant, agreed with Ms. Leary about the usefulness but also the impracticality of mandating counting animal numbers. He wondered if there was a funding avenue that could be applied to this need, such as a grant incentive. There might also be an opportunity to give recognition to institutions that have made progress in this area. Fostering collaborations could provide opportunities for organizations to share best practices and learnings. Information about the efforts to advance use of NAMs and replace animal use needs to get out into the public, outside of scientific journals and websites that have limited viewers. Public support will help drive progress in this area. The tables that were shown in these presentations tracking animal reduction were powerful examples of progress, but it would be of interest to know where in the development process the animal reductions were realized. It's also important to identify what species are being impacted. Finally, Dr. Baines stressed that current knowledge about alternatives to animal use needs to be incorporated into educational pipelines, and interactions with academia would help with this.

Discussants for the subtopic of "Industry Approaches" were asked to consider the following questions:

- Which options and approaches used by other groups (e.g., private industry, international organizations, etc.) could ICCVAM agencies adapt in reporting metrics?
- What stakeholder organizations are in the best position to assist in collecting information on animal use and implementation of NAMs, and why?

Dr. Gehen, first discussant, appreciated the level of specificity that was achieved in the presentations given today. Measuring progress in this area is complicated but necessary to make sure we are moving in the right direction. Dr. Marty's presentation provided some practical ideas that have the potential to be applied by others. It's important to understand what we are working toward and ensure that the metrics we are using are truly meaningful. One potential approach to establish an appropriate denominator could be measuring number of animals per year, but another equally valid approach might be animals per molecule or submission. Understanding the impact of variability is also important. Any baseline that is established needs to represent cumulative data to be realistic. While it's intuitive to expect that more NAMs being approved will have an impact on animal use, linking that use to animal savings is important to show progress. Comparing the number of animals used to the test guideline requirements raises questions that need to be examined; it could be that studies to address several needs are being combined in a single study. Regarding what stakeholder organizations are best positioned to obtain these data, Dr. Gehen felt that both contract research organizations and regulatory agencies could play important roles. He appreciated EPA's efforts to assess and understand the impact of waivers but noted that such analyses do not include products that fail to make it to registration, and consideration needs to be given to how to capture that information.

Dr. Sura, second discussant, felt that companies that are dedicated to the 3Rs and advancing NAMs should be interested in measuring progress in reducing animal use. That requires defining the context in which NAMs are being used. It also needs to be recognized that while NAMs are being accepted more broadly, there is a feeling in certain sectors that animal testing is still needed. She echoed Dr. Gehen's recognition of EPA's efforts, in particular those of HASPOC, and she noted OPP's approachability and transparency. She felt that there might be an opportunity to involve trade associations in raising awareness of NAMs and dispelling the idea that animal tests are the only option and suggested trade shows as a potential venue through which to accomplish this.

Discussants for the subtopic of "Workgroups" were asked to consider the following questions:

- What additional requirements in the IACUC review process could be considered to increase the identification and consideration of NAMs?
- What approaches could be considered to raise awareness of NAMs and thereby lead to their consideration and use by researchers and/or regulators?

Dr. Berg, first discussant, commented that additional requirements to consider alternatives in IACUC reviews would be great. Most scientists are trained by principal investigators who are not knowledgeable about alternatives, so a training requirement might be helpful. Barriers to implementation of NAMs include lack of knowledge, lack of confidence, and lack of access. Training could overcome many of these barriers; addressing lack of confidence requires appropriate data sets connected to human outcomes. There is still a lack of awareness about how poorly animal studies predict human health effects, and this will only be addressed by availability of such data. She echoed the disappointment expressed by others in the FDA's proposal to run more animal studies for the purpose of validating NAMs, emphasizing that benchmarking against human data is a better approach. On the other hand, she praised FDA for their

work on developing standards for in vitro pharmacology data, which will enable harmonization of data and creation of data sets that will support future predictive studies. To address access to alternatives, she suggested supporting use of alternatives through granting mechanisms, which would have the added benefit of helping build the data sets that will support greater confidence.

Dr. Clippinger, second discussant, mentioned that IACUC members should have both an initial requirement and ongoing training in alternatives. She noted the important role that a centralized reviewing body can play in reducing duplicative testing. It would also be useful to have individual subject matter experts review proposals and double-check for opportunities for alternatives use and opportunities for collaboration. Foundational to all of this is a change in mindset; people don't consider the IACUC review meaningful. Activities that could help change that include education on the benefits of non-animal testing and tracking and identifying areas of high animal use, with a focus on frequently used tests. Competition and recognition could stimulate this process. Regarding raising awareness of NAMs, groups such as ICCVAM need to seek out opportunities to increase the diversity of interactions, including interacting with communities with whom they may not be comfortable. Other avenues to pursue might include translating recorded presentations into other languages, putting requirements about alternatives in graduate course work, and mentoring younger scientists through the Society of Toxicology or other organizations. Providing incentives is a great idea, and surveys can help identify key gaps. While it's important to publish about new methods and case studies in scientific journals, this is not a substitute for clear messaging from agencies about their information requirements. She closed by identifying NICEATM News as a good information resource and encouraged all on the call to subscribe and contribute[24].

### *Additional SACATM Comments*

Dr. Tal noted that the EPA tables were easy to find online, but she agreed with Dr. Gehen that more granular information about the phases of testing in which animals are used would be helpful. Using programs already in place is key to education and promotion. While it's great to see agencies providing more information, it would be helpful to consolidate it, perhaps by developing a dashboard on efforts to develop NAMs. Small funding programs could help offset costs of NAMs that are expensive because of proprietary technology.

Dr. Page applauded the steps that have been taken to measure progress and appreciated this being identified as a focus point of this meeting. She especially welcomed the ideas on implementing qualitative measures and stressed the importance of opportunities to compare ideas and practices. She liked the points made by Dr. Ortiz Franyuti about data storage and transfer; these activities are key to developing strong informative metrics. She wondered whether Roche's internal metrics might take into account the number of animals used to get a molecule to the clinic. While the Dow publication doesn't share specific animal numbers, it is useful because sharing their approach can help others develop their own metrics. The number of regulatory submissions is important to track to provide context to animal use. She wondered whether the proposals suggested by Dr. Rattner for use within DOI could be leveraged

[24] Subscribe at https://list.nih.gov/cgi-bin/wa.exe?A0=NICEATM-L&X=CA8F8490FB6AD4644F&Y; recent articles available at https://ntp.niehs.nih.gov/go/niceatm-news.

to provide an approach to providing feedback to industry. Prerecorded presentations could be posted on agency websites to provide information while guidance is being developed. To facilitate data sharing, the issue of maintaining confidentiality needs to be addressed. It's important for agencies to educate registrants about available alternatives for specific endpoints when they see registrants continuing to submit animal data for those endpoints. She acknowledged the challenge faced by agencies that don't receive submissions, such as CPSC. Use of animal numbers and metrics make things more complicated, and she appreciated the efforts to broaden the types of approaches that can be used to measure progress, as represented by today's presentations. She closed by reiterating points made earlier about the importance of education of the current workforce and of collaboration among a diversity of groups.

Dr. Nañez felt that Dr. Marty presented a good example of how companies are taking an active approach to reducing animal use. A session on this topic should be organized for a future Society of Toxicology meeting. He felt that Mr. Manuppello's oral comments described an activity (microsampling) that could be done quickly with the potential to make a big impact. Taking microsamples for toxicokinetics from mice that are already being treated for other endpoints would reduce animal use and enable better parallels to be drawn between the toxicokinetics and other effects, and he suggested that FDA could encourage this practice.

Dr. Ortiz Franyuti, responding to Dr. Page's comments about data management, wondered how industry might establish standards for this and suggested that controlled terminology would be a key element. Good data management is needed so these data can be used far into the future as technologies and knowledge progresses. Responding to Dr. Berg's comments about supporting in vitro methods use, she suggested that regulatory agencies request data on animal use or use of in vitro methods in preclinical testing.

Dr. Denis Fourches, Oerth Bio, agreed with previous commenters that more granular information about the phases of testing in which animals are used would be helpful in the context of opportunities for waivers. He noted that the utility of NAMs, or at least their established applicability, can be limited for certain chemical types. He agreed with points that had been raised about protecting intellectual property and patient privacy but encouraged industry to explore how to share data while addressing these concerns.

Dr. Berg noted that FDA has an ongoing collaboration to create a framework of templates and data standards for in vitro data.

Dr. Charest suggested that industry consortia exist that could play a role in facilitating data collection, as could organizations that gather data for industry use. He suggested there might be value in examining rates of success of new products in the context of NAMs use. Evidence that using NAMs to test products improves success rates would be a strong incentive for their use.

Dr. Szczepan Baran, VeriSIM Life, emphasized the value of clarifying and harmonizing terminology such as "fit-for-purpose", "validation", etc. Discussions of benchmarking need to include guidance for improvement. He suggested that initiatives going on in Europe such as the Innovative Medicines Initiative[25] represent an example of how a

[25] https://3tr-imi.eu/about/vision-and-objectives.

range of institutions can work together to share data.

Dr. Marty agreed with Dr. Gehen's point that the people who are most involved in using NAMs are the best ones to help build confidence in their use. She stressed the importance of sharing information about compounds that have failed in development to highlight how NAMs can be used to identify compounds that should not advance. Companies should share more NAM data with regulators, and regulators are going to need the resources to review these data sets and provide feedback on whether they are fit-for-purpose and if not, why not. She suggested that it might be useful to require subject matter experts on NAMs to serve on IACUCs, not only to inform on available NAMs but to gather information on animal tests that are still being done due to lack of available NAMs as an opportunity for NAMs development.

Dr. Kleinstreuer noted the relationship between the topic of metrics and validation; today's presentations and discussion will provide a good background for tomorrow's program on that topic. She mentioned publications that have come out recently about quality standards for NAMs, including the Johns Hopkins-led effort to develop Good Cell Culture Practice 2.0[26] (GCCP) and the OECD Guidance Document on Good In Vitro Methods Practices[27] (GIVIMP), as well as an effort being led by the United Kingdom National Centre for the 3Rs to develop standards for in vitro methods data reporting. She agreed with the need for harmonized terminologies and standardized ontologies; these are key to implementing FAIR standards. Finally, Dr. Kleinstreuer stated that we need to fully leverage the use of human data to achieve biological relevance; she asked SACATM members to think about sources for these data for tomorrow's discussion.

Dr. De Abrew thanked the day's presenters and discussants and adjourned the meeting for the day at 3:05 p.m.

# September 22, 2022

Dr. De Abrew called the second day of the meeting to order at 10:01 a.m. SACATM members and key NIEHS staff introduced themselves. Dr. Brownlow reviewed meeting logistics and read the conflict-of-interest statement.

## VIII. Validation and Establishing Scientific Confidence in NAMs

### ICCVAM Validation Workgroup: Updating the ICCVAM Guidance on Validation – Progress Report

Dr. Suzanne Fitzpatrick, FDA, provided an overview of the activities of the ICCVAM Validation Workgroup, which has representation from ten ICCVAM agencies. The workgroup is updating the ICCVAM document "Validation and Regulatory Acceptance of Toxicological Test Methods," which was published in 1997.[28] While much of the information in the document remains relevant, some of it is outdated. Importantly, the 1997 document neither reflects the high level of collaboration taking place today, nor

[26] Available at https://doi.org/10.14573/altex.2111011.
[27] Available at https://www.oecd.org/chemicalsafety/guidance-document-on-good-in-vitro-method-practices-givimp-9789264304796-en.htm.
[28] Available at https://ntp.niehs.nih.gov/iccvam/docs/about_docs/validate.pdf.

considers the "context of use" concept. In addition to context of use, other key concepts the workgroup is focusing on in this update include:

- Technical characterization.
- Information transparency.
- Data integrity.
- Biological relevance.
- Independent review.

Expanding on the idea of context of use, Dr. Fitzpatrick explained that the end use of the method should determine the level of validation needed for it. For example, the validation approach for a method intended as a screen will differ from one intended as a full replacement of a method addressing a regulatory requirement. The important question is what the consequence of a wrong answer could be. The new guidance will foster the use of flexible, efficient, and robust practices to establish confidence in new methods.

Another key topic that will be addressed in the new guidance is relevance: biological relevance, biological plausibility, and mechanistic relevance. The guidance will also discuss quality of reference data, the role of legacy animal data, best practice for quality systems, sources of variability, standards for qualification or validation, and incorporation of data quality tools. The workgroup will consider how the principles articulated in the new document fit into a globally harmonized approach to support continued mutual acceptance of data and working with international partners to achieve that. The new document will reference established and accepted documents such as OECD Guidance Document 34. The role of ICCVAM in validation includes assuring an independent validation process, advising federal agencies on validation strategies, facilitating collaborations, and encouraging communications.

Dr. Fitzpatrick concluded by summarizing next steps. The Validation Workgroup is organizing and finalizing the document while incorporating input from ICCVAM agencies. The draft document will be reviewed by ICCVAM agencies before being released to the public for comment. Dr. Fitzpatrick emphasized that validation and qualification of new methods is ongoing by the individual agencies even as this document is being developed.

***Clarifying questions and comments***: Dr. Berg asked Dr. Fitzpatrick when she expected the draft document would be made available for public comment, and Dr. Fitzpatrick expressed the hope that that would happen later in 2022.

## Technical Framework for Enabling High-Quality Measurements in NAMs

Dr. Elijah Petersen, National Institute of Standards and Technology, noted that his presentation was based on a paper recently published in ALTEX[29] with coauthors from CPSC, the U.S. Department of Defense, and NICEATM. This paper described a framework for evaluating quality; such a framework could be applied to a scientifically relevant NAM that has the potential to fill a testing need. The first part of this framework describes a set of basic quality tools that could be applied to build confidence in NAMs.

[29] Available at https://doi.org/10.14573/altex.2205081.

Applying these tools is a two-step process involving a conceptual evaluation of sources of technical variability, done before going into the lab, and then an evaluation of assay performance that involves performing the assay in the laboratory. Once data are obtained, the third step in the framework is applied to add statistical confidence to decisions based on NAM results and to understand the cumulative uncertainty. Transferability might then be evaluated, but this is situation-specific and may not be needed for all methods. These steps are all interrelated. An earlier paper (Rosslein et al. 2015[30]) described in detail an approach to apply cause-and-effect analysis that examines and classifies several sources of variability in a cytotoxicity assay to evaluate engineered nanomaterials. Control measurements should be present in each step of the assay protocol. Dr. Petersen used a plate design as an example of how to incorporate appropriate controls to identify variations in elements such as pipetting or test substance interference. Control charting can identify sources of variability that arise over time. Scatter plots can identify interactions among variables. Histograms can identify variations in data distribution that may be calling attention to issues with the assay and the distribution of the data obtained (e.g., normally distributed). An example from a skin sensitization test illustrated use of a statistical evaluation that could be applied to consider assay-to-assay uncertainty when making assay calls. In summary, he noted this framework reflects perspectives from different agencies to develop a user-friendly approach to facilitating standardization and adoption of NAMs.

***Clarifying questions and comments:*** There were no clarifying questions.

## Scientific Confidence Framework: Biological Relevance – a Better Benchmark

Dr. Kleinstreuer began her talk by introducing a framework for flexible, fit-for-purpose NAMs validation. One of the elements of this framework is biological relevance, the idea that relevance to the biology of the species of interest should be an important consideration when validating a NAM. Biological relevance should be supported by mechanistic evidence. An important piece of mechanistic evidence is an adverse outcome pathway (AOP) for the endpoint of interest, which serves as an organizing framework for interpreting chemical effects. Her presentation provided three examples where human biological relevance is being considered in the context of establishing confidence in NAMs for regulatory use.

One set of projects used human biology as the basis for validating existing methods for assessing topical toxicity. In the first project, a retrospective analysis of dermal absorption data found that a human cell-based system provided a more conservative estimate of dermal absorption than the standard approach, which considers in vivo rat and in vitro rat and human data. A second project focused on AOP-based defined approaches for skin sensitization used human data for the basis of comparison to a standard mouse assay. In all cases, the defined approach based on human biology outperformed the animal data in predicting the human data-based hazard classifications and potency estimates. Finally, existing human-relevant approaches for assessing eye irritation potential address concerns that the rabbit model might have limitations for predicting human effects. A recent paper by NICEATM, EPA, and collaborators[31] considered human anatomy and which in vitro methods can best measure mechanistic

[30] Available at https://doi.org/10.1021/tx500327y.
[31] Available at https://doi.org/10.1080/15569527.2021.1910291.

effects in the human eye.

The second group of projects represented efforts to leverage understanding of human disease mechanisms and to design screening batteries that can inform potential chemical effects for longer-term toxicities. A project that used a human stem cell assay and IVIVE to measure developmental toxicity potential for valproic acid analogs indicated that this approach may be more protective than animal assays. Another project examined DNT assays with respect to their human relevance and combined a group of the most human-relevant assays into a test battery. An OECD case study publication describes the use of this battery to prioritize in vivo testing of organophosphorus flame retardants.[32] NICEATM is also using computational methods and high-throughput screening data in an approach to better understand the role of environmental exposures in cardiovascular disease, an endpoint not usually considered in regulatory testing. Knowledge about human cardiovascular failure modes was applied to data from relevant human cell-based assays to create cardiotoxicity bioactivity profiles for a group of chemicals. This identified cellular concentrations that could cause cardiotoxicity that can then be related to external exposure levels to prioritize further testing. This study confirmed some known cardiotoxins and identified potential new ones.

The final group of projects used complex human biology-based platforms, specifically microphysiological systems, to provide insights where animal models might not be sufficient. NICEATM, along with two other NIH institutes, the U.S. Department of Defense, and other collaborators established the Microphysiological Systems for COVID Research Working Group to gain insights into COVID-19 effects for which there aren't good animal models. One model developed at the National Institute of Allergy and Infectious Diseases is being used to test candidate drugs. Results have correlated well with human clinical results, and the model is being used to support repurposing of existing antivirals and other drugs in advance of clinical trials.

In the spirit of improving human relevance of NAMs technologies, Dr. Kleinstreuer noted that NICEATM will be presenting a symposium webinar[33] on using NAMs to address variability and susceptibility across populations.

***Clarifying questions and comments:*** Dr. Tal asked Dr. Kleinstreuer to reflect on the process for achieving acceptance of the defined approach for skin sensitization[34]. Dr. Kleinstreuer responded that it was important to highlight the shortcomings of the standard animal tests, even though they were widely recognized as being adequately protective. Being inclusive and collaborative to establish trust was also essential. She added that since the OECD skin sensitization guideline was adopted, a new guideline for defined approaches for assessing eye irritation potential was adopted much more quickly. Dr. Charest noted that in the dermal absorption project the human model was a more stringent test than the rat model, and he asked Dr. Kleinstreuer to comment on a situation where the human model might be less stringent. Dr. Kleinstreuer agreed that this kind of situation presents a challenge. In the case of the skin sensitization defined

[32] Available at http://www.oecd.org/officialdocuments/displaydocument/?cote=env/cbc/mono(2022)26&doclanguage=en.
[33] Information at https://ntp.niehs.nih.gov/go/popvar.
[34] Guideline 497; available at https://www.oecd-ilibrary.org/environment/guideline-no-497-defined-approaches-on-skin-sensitisation_b92879a4-en.

approach, the standard mouse assay was found to be overly sensitive with respect to human biology, producing a lot of false positives when compared to human data. The inclination to regulate according to the more conservative test was part of the difficulty with getting the defined approach guideline adopted. NICEATM is currently trying to determine the mechanistic basis for the false positives in the mouse test.

## Variability of Reference Data

Dr. Agnes Karmaus, Inotiv (contractor supporting NICEATM), began her talk by noting that there are a lot of animal data available from guideline studies. NICEATM and partners have examined the reproducibility of data from animal studies assessing eye irritation, skin irritation, and acute oral toxicity, with the goal of providing appropriate context for validation of NAMs. Dr. Karmaus' presentation focused on reproducibility of categorical assignments for hazard classifications. Categorical classifications for some endpoints can be based on observation rather than quantitative measurements. NICEATM evaluated reproducibility by computing conditional probabilities which consider the resulting hazard classification from a specific study and calculating the probability that a subsequent study would replicate that classification.

Applying this approach to a data set of 500 substances tested at least twice in the rabbit eye test revealed that the least and most severe irritation categories were the most reproducible. However, categories identifying mild and moderate irritants had lower rates of reproducibility. Similar results were found in an analysis of the rabbit skin test.

NICEATM's analysis of rat acute oral toxicity used classifications based on LD50 data for thousands of chemicals. In this case, reproducibility was greatly impacted by the LD50 range of each category; it's harder to reproduce a classification within a category as the range of LD50 values represented by that category gets smaller. For some categories, the probability of a classification being reproduced by a subsequent test was near or below 50%. The database was very well-curated; Dr. Karmaus stressed the importance of having a high degree of confidence in the data sets used for these analyses. To define a margin of uncertainty, NICEATM used curated point estimate LD50 values and bootstrapping to define a range that encompasses most experimental LD50 results.

In summary, characterizing variability in reference data can help provide context for existing guideline studies, set reasonable expectations for NAMs reproducibility, and define a margin of uncertainty to apply to in silico predictions and alternative methods.

***Clarifying questions and comments:*** Dr. De Abrew asked Dr. Karmaus to clarify the procedure used for the conditional probability analysis and whether any factors had been identified that affected reproducibility. Dr. Karmaus responded that physicochemical properties, use categories, and other chemical properties were examined but none had clear correlations with reproducibility. In response to a question posed by Dr. Page, Dr. Karmaus replied that a lot of the data were from limit tests; these weren't removed from the initial analysis, but the data set required heavy curation to remove duplicates. The limit test data were excluded from the data set used to estimate confidence intervals. Dr. Page then asked how these analyses handle endpoints such as corrosion. Dr. Karmaus acknowledged that deriving an uncertainty margin from categorical data is challenging but NICEATM is currently considering how to do that.

## Transparency, Data Integrity, and External Review

Dr. João Barroso, European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM), began his presentation by reminding the audience of the elements of the framework for establishing scientific confidence in NAMs: fitness-for-purpose, human biological relevance, technical characterization, independent review, and data integrity and transparency. The last two elements were the focus of his presentation.

The modern safety assessment toolbox includes existing animal data, in vitro testing to measure cellular or genomic activity, in silico approaches that can use machine learning or artificial intelligence to predict toxicity, and human biokinetics and biomonitoring. Some of these methods have proprietary elements that are confidential or complex, resulting in lack of transparency. OECD member countries have agreed that test guidelines should not contain elements that are confidential to an extent that hinders adequate scientific validation and independent review, and OECD has put practices into place that make this possible without compromising intellectual property. Independent scientific review is an important part of building confidence in a new method, and especially important for new technologies or applications. Independent review activities should not be managed by method developers. EURL ECVAM has a process for method review through its Scientific Advisory Committee. This process was most recently applied to the GARD®skin genomic skin sensitization assay, an innovative technology that uses transcriptomics and an independently reviewed machine learning algorithm. All proprietary data were made available to the scientific advisory committee during the review, which increased trust and enabled regulatory acceptance. He noted that the EURL ECVAM Tracking System on Alternative Methods[35] (TSAR) database makes available information about review and regulatory acceptance of methods.

Data integrity is an important element of validation. Ideally, studies should be conducted under Good Laboratory Practice (GLP) conditions but, in the past, noncompliance with GLP was not considered a problem because studies were coordinated by independent parties such as EURL ECVAM who guaranteed the independence and integrity of the study. When validation is managed by commercial entities, these are called into question. It's also becoming harder to identify inconsistencies or fraud in high-content omics data sets or machine learning algorithms. Therefore, GLP compliance or independent quality assurance is necessary to establish integrity and credibility. Standards for these are described by the OECD GIVIMP document.

Demonstrating reproducibility of a method is essential, particularly within a laboratory. Ring trials are the traditional means by which interlaboratory reproducibility is established. These are very expensive, and their outcomes can be affected profoundly by laboratory quality or expertise; proper training and transfer studies are essential. Better characterization of critical steps by the developer and sensitivity analysis of parameters that can affect outcomes can be at least as important for optimizing reproducibility as conducting a ring trial. Dr. Barroso described a validation study EURL ECVAM is currently conducting on a method to assess thyroid activity as an example of a more efficient approach than a ring trial.

[35] Available at https://tsar.jrc.ec.europa.eu/.

In closing, Dr. Barroso emphasized the need for validation to keep pace with rapid scientific progress while still maintaining scientific integrity and usefulness. He also stressed that validation is not the same as regulatory acceptance.

***Clarifying questions and comments:*** In response to a comment by Dr. Tal that there are a number of methods described as halted or stopped on the TSAR website with no additional information provided, Dr. Barroso indicated that information about method validation is not made public until peer review. If a validation study is ended before peer review, details about the study are kept confidential to allow the developer to use that information to improve the method. Most often, the method doesn't move forward because it lacks a clear fitness-for-purpose or context of use. Clarity around regulatory information requirements would make validation easier. Responding to a second question from Dr. Tal, Dr. Barroso replied that the thyroid method under evaluation used a battery approach. A group of methods is being assessed in parallel with the same chemicals and evaluated together to determine how they complement each other and inform the adverse outcomes. Dr. Tal then asked if EURL ECVAM considered calling for developers to address a specific domain such as DNT. Dr. Barroso said that has been done for clearance, and EURL ECVAM is working on a number of methods for DNT in collaboration with other parties. These aren't in TSAR because they weren't submitted specifically to EURL ECVAM. Dr. Charest asked if there are any contexts in which use of the GIVIMP standards is required. Dr. Barroso responded no, and specified that GIVIMP standards are not required under the Mutual Acceptance of Data agreement. One advantage of GIVIMP is that it addresses a gap in European Union requirements, which do not require that nonclinical safety assessments be done under GLP, and that also applies to validation studies for in vitro methods. Efforts are underway to develop certification programs utilizing GIVIMP for method development and validation. Responding to a question from Dr. Charest, Dr. Barroso replied that GIVIMP standards are not stricter than GLP, but they're complementary by focusing on quality to a greater extent than GLP, which focuses more on transparency.

## Public Comments

One written public comment was submitted for this section on behalf of HSUS/HSLF.

### *Oral Public Comments*

Ms. Haugen expressed concern about biases being applied to information made available to the public and the need to be aware of people's inclination to resist information that they do not agree with. Most clinicians are not properly trained in diagnosing toxicity and are not familiar with how to treat cases of chronic toxicity. In particular, they may not consider toxicity as a cause of illness if the product to which the patient was exposed is considered nontoxic. There is a sense among clinicians that a "nontoxic" label on a product can be relied upon for diagnostic purposes. The lack of a mandatory reporting requirement for toxicities further limits the availability of information about these cases. Such information could be used along with human-relevant NAMs to capture real-world impacts of chemical harm and better protect children and other vulnerable groups. However, relying on animal data for NAMs validation will continue to perpetuate inaccurate perceptions of chemical effects in humans.

Ms. Kristie Sullivan, representing PCRM, agreed with Ms. Haugen on the need for

greater monitoring of toxicity and training of clinicians to recognize it. She expressed concern about the length of time it took to make notable progress in developing NAMs for acute endpoints and encouraged action that would enable faster adoption of NAMs for chronic endpoints. Adoption of new approaches to validation and building confidence in NAMs will be critical to this. In turn, central to these is agreement among developers, users, and regulators about context of use and regulatory need. Test method developers need to clearly understand how regulators make decisions and the role data play in those decisions. She encouraged efforts to use more human reference data, and to make sure these data are standardized. Mechanistic relevance is also important, and benchmarking models or approaches to appropriate mechanistic pathways. Data generation needs to be driven by physical, chemical, or other mechanistic properties, and not industrial or regulatory sector. Likewise, models should be evaluated by collaborative peer reviews that meet the needs on many agencies. To improve confidence within the community at large, she encouraged ICCVAM to continue to work with OECD and the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Small companies and ingredient suppliers represent a gap in engagement, and NICEATM and ICCVAM should explore ways to engage these stakeholders. Companies also need more training on how to use NAMs data. Finally, agencies should consider requiring NAMs use in certain contexts.

### *Comments from Designated SACATM Discussants*

Discussants for “Scientific Confidence Framework” were asked to consider the following questions:

- How can the qualitative and quantitative variability of in vivo reference test methods be best applied when evaluating the performance of NAMs?
- How should biological relevance, both of in vivo reference test methods and of NAMs, be considered in establishing scientific confidence in NAMs?
- How can all of these aspects be incorporated into a scientific confidence framework to ensure that NAMs performance is as good or better than the animal tests they are intended to replace?

Dr. Tal, first discussant, noted that the Scientific Roadmap states the importance of consideration of human data for validation of NAMs. NICEATM has done a good job of demonstrating the variability of animal data, calling into question the use of these data as an anchor point and in particular the need to generate new animal data to validate NAMs. A lot of effort has been put into NAMs technical characterization to ensure they generate high-quality reproducible data, without the expectation to exceed the reproducibility of the animal test. This is also true for limits of detection. Meeting these standards and demonstrating biological relevance should be sufficient for acceptance. Dr. Tal expressed interest in seeing case studies on more complex NAM test batteries, particularly in DNT. Regarding biological relevance of NAMs, while less complex than animal models, they are mostly human-based, which can improve relevance, especially when tied to valid AOPs. A few practices that can improve confidence include using the same positive controls as in the animal test, establishing and using lists of reference chemicals, and use of gene editing and pharmacological manipulations to establish the

validity of key events. Causality linked to an adverse outcome would provide strong weight of evidence for a NAM assessment within a scientific framework.

Dr. Page, second discussant, commented that an updated ICCVAM guidance document will be a major milestone in progress. She reiterated the importance of relating NAMs validity to human health rather than reproducing animal data, although she acknowledged that some animal data can be useful, especially in those cases where human data are not available. Clarity of purpose and applicability domain is important to achieve validation of methods, and Dr. Page agreed with the concept of not limiting NAMs validation to particular sectors; it's more fruitful to define hazard categories based on mechanism of action. All the work on variability is valuable; alternatives should not be expected to be better than the animal test. On the other hand, control generation needs to be balanced with assay cost; incorporation of a lot of controls could make an assay prohibitively expensive. She agreed with points that had been made about alignment with AOPs being important to establishing the biological relevance of NAMs. The Clippinger et al. paper is a good example of how to compare animal and human pathways in the context of validation. Similarly, Dr. Petersen made good points about developing a validation framework that ensures that key points are covered during the method development process. However, this consensus needs to be reached globally to ensure that requirements for animal data are not used as an excuse for not using NAMs.

### *Additional SACATM Comments*

Dr. Ushio noted that biological relevance is best supported by knowledge of the underlying biology, which provides confidence that the reason for a chemical effect is clearly understood.

Dr. Fourches commented that many quantitative structure–activity relationship (QSAR) models used today do not account for variability of the training data, and this needs to be understood when they are used. In response to Dr. Fourches' comment, Dr. Kleinstreuer said that NICEATM considers this point in their computational crowdsourcing projects, and consensus models like the Collaborative Acute Toxicity Modeling Suite (CATMoS) provide a confidence interval that takes that into account. She described a situation where the model was used to identify an error in the training data.

Dr. De Abrew commented on the importance of having well-established reference compounds for building confidence in NAMs. He asked Dr. Barroso, in the context of European classification and labeling requirements, whether EURL ECVAM developed NAMs for a specific function. Dr. Barroso replied that EURL ECVAM's primary focus is assessing and testing new technologies rather than developing NAMs. However, when EURL ECVAM interacts with method developers, it likes to see methods that inform dose-response and points of departure. Of course, in the regulatory context, those results need to be translated into categorical classifications. There are many different regulatory sectors within the European Union. For some of these, risk assessments are important, so methods are needed that provide more than just a categorial classification. Dr. De Abrew asked Dr. Barroso to comment on how that fits into supporting broader acceptance of NAMs. Dr. Barroso responded that a bigger problem than translating dose-response information is incorporation of exposure information into risk assessment.

Dr. Charest agreed with the importance of consideration of biological relevance and noted that such consideration needs to take normal function into account and ensure that representation of normal function is consistent across models.

## Understanding Context of Use for Medical Devices: Case Study – U.S. Food and Drug Administration

Dr. Shelby Skoog, FDA Center for Devices and Radiological Health (CDRH), reviewed the FDA's 2020 biocompatibility guidance and other FDA avenues for consideration of alternate methods. The CDRH Medical Device Development Tools (MDDT) program is a voluntary program through which developers can establish the context of use for new safety, effectiveness, and/or performance evaluation tools for medical device development. One type of MDDT is a nonclinical assessment model that measures or predicts device function or in vivo performance. Such models can include in vitro models to replace animal testing. Dr. Skoog reviewed the concept of context of use and noted that an MDDT is qualified for a particular context of use, although context of use for an MDDT can be adjusted based on the qualification data provided. Considerations for NAMs to assess biocompatibility include the endpoint(s) being considered and whether a specific biocompatibility (or multiple tests) test is being proposed for replacement. For example, considerations for skin irritation would be different for limited contact with intact skin, repeated contact, or intracutaneous use. Another consideration would be the mechanism of action or biological endpoint being considered by the traditional test as compared to the NAM; in other words, how screening with the proposed NAM would inform the relevant outcome.

Dr. Skoog reviewed types of qualification data, including scientific literature and stakeholder data, and noted that FDA encourages use of existing data to support NAMs qualification. The relevance of the existing data is considered, as well as what data gaps need to be filled. As an example, she reviewed the applicability and limitations of tests described in OECD Test Guideline 439 for skin irritation, which is validated for classification of neat chemicals as Category 2 irritants but may not be used to definitively classify a chemical as a Category 3 mild irritant or for testing of mixtures or specific hard-to-test chemicals. Testing of medical devices requires a particular applicability domain: the approach needs to support testing of chemicals with a range of potencies and diverse properties, and support testing of dilute concentrations and mixtures. Test protocols may need to be modified to be suitable for both polar and nonpolar extracts, consider exposure duration, and other factors. The NAM needs to be demonstrated to clearly distinguish between positive and negative responses, and there needs to be justification that its performance is adequate for the proposed context of use and how it performs in comparison to any available in vivo data. She closed by showing examples of types of data that are requested, questions that might be asked of the developers, and information about resources available to alternative methods developers.

***Clarifying questions and comments:*** There were no clarifying questions.

## Context of Use of Mammalian Median Lethal Dose in Ecological Assessment at U.S. Environmental Protection Agency

Dr. William Eckel, EPA OPP, presented work in progress toward evaluating the potential role of the CATMoS QSAR model in supporting testing waivers for acute oral toxicity. He

reviewed the EPA OPP's role in developing risk assessments for pesticides, which consider both risk to humans handing the pesticides and wildlife that might be exposed to them. For wildlife risk assessment, EPA's Terrestrial Risk Assessment (T-REX) tool uses studies of residues on specific foraging items to estimate exposures to small, medium, and large animals, then derives a risk quotient based on adjustments to the rat LD50. In the current study, CATMoS predictions of rat oral acute LD50 were compared with OPP in vivo data for 178 active ingredients. Preliminary data indicate that CATMoS predictions of LD50s greater than 2000 mg/kg are reliable; this is a common limit dose for substances expected to be nontoxic. This could be used to determine a threshold application rate. These results support potentially using CATMoS predictions in lieu of in vivo testing in some cases, depending on the LD50 prediction and the proposed use of the pesticide.

***Clarifying questions and comments:*** In response to a question from Dr. Fourches, Dr. Eckel confirmed that all chemicals considered in this study were within the CATMoS applicability domain. Dr. Charest asked whether the T-REX tool assumes uniform application of a pesticide or accounts in any way for variability across the application space. Dr. Eckel replied that while the only environmental fate factor T-REX accounts for is decay of the pesticide over time, which can be adjusted based on the specific properties of the pesticide, it does consider multiple applications. Other factors such as geography are incorporated into the subsequent extended risk assessment. Dr. Nañez asked if the evaluation of the specific 178 active ingredients included other molecules that may have been considered in the development of these chemicals. Dr. Eckel indicated that the study was limited to EPA data on active ingredients registered with the agency in the last 20 years. Responding to a second question from Dr. Nañez, Dr. Eckel commented that decisions on waivers are made within OPP, and guidance documents that explain the rationale for those decisions are available online.

## Development of a Rapid Risk Assessment Process and Software Tools to Support Air Force Operational Decision-making and Technology Acquisition

Dr. Andrew Keebaugh, U.S. Air Force Research Laboratory (AFRL), presenting on behalf of Dr. Rebecca Clewell, described a process the laboratory is developing to provide rapid assessment of chemical risk to aid decision-making within Air Force operations. NAMs are of interest in this context because they can provide information faster than animal studies. The U.S. Department of Defense does not have a regulatory role but the data they generate can be incorporated into regulatory risk assessments.

The Predictive Risk Capability Build is being used internally as a tool to incorporate NAMs into evaluations of emerging hazards to Air Force personnel and uses a workflow to arrive at either a risk estimate and safe exposure estimate or a determination that more information is needed. The process varies for chemicals with existing data versus novel chemicals. Evaluations incorporate several elements including in vitro tests and QSAR and PBPK models. Key areas of focus are respiratory toxicity and neurotoxicity. Dr. Keebaugh reviewed results from a novel artificial intelligence model that predicts acute toxicity based on GHS classification and chemical structure, and another model that predicts ligand binding to neurotransmitter receptors to identify potential neurotoxins. Another function of the Predictive Risk Capability Build aggregates data from a number of databases to establish a hazard index. An inhalation toxicity model for poorly

characterized chemicals is based on the AOP for lung fibrosis.

AFRL is also developing customer-facing products, including the ToxAdvisor desktop app and the ToxAdvisorLite mobile app. AFRL envisions that these apps will provide information on hazard and personal protective equipment requirements. Features will indicate if a hazard assessment is based on experimental or predicted data, provide comparisons to well-known benchmark chemicals, and color code for easy interpretation by a broad audience.

***Clarifying questions and comments:*** Dr. Nañez asked if the public-facing tools described will be made available outside the Air Force, and Dr. Keebaugh replied that they will only be used within the Air Force.

## Public Comments

### *Comments from Designated SACATM Discussants*

Discussants for "Understanding Context of Use" were asked to consider the following questions:

- What would you suggest to best support successful harmonization, both among U.S. federal agencies and internationally, in the most efficient and effective way?
- What are the most important components of a scientific confidence framework specific to each of the contexts of use presented by ICCVAM agencies?

Dr. Nañez, first discussant, commented that strong guidance from agencies would play an important role in harmonization. It would also be good to have more crosstalk among agencies about effective tools available and cited the Air Force's ToxAdvisor app as an example. Agencies should also discuss with industry how to apply NAMs to reduce animal numbers or make the risk assessment more predictive. He cited Mr. Manuppello's public comment as an example of the kind of data that could impact animal numbers; positive feedback from agencies could encourage further similar actions. EPA is taking actions to implement waivers and support other approaches to reducing animal use that could be used in other sectors, for example for prioritization in early stages of development. There is a need to exploit data within industry silos to facilitate these kinds of actions.

Dr. Charest, second discussant, stated that harmonization needs to focus on biological relevance, in particular to get agreement across sectors about what is biologically relevant. It would also be helpful to identify contexts of use that can be applied across sectors. On the other hand, test method developers should not avoid contexts of use being very specific. Agreement across agencies on these specific contexts of use will maximize the impact of methods qualified to address them. He suggested the creation of a registry to track NAMs having specific contexts of use that would enable connections to be made among complementary methods. Dr. Charest acknowledged that this would be difficult to do in the context of medical devices, which have a lot of unique issues. Regarding a scientific confidence framework, biological relevance is also key here. Human-based NAMs have an advantage here and might help us get away from reliance on animal data. He emphasized the importance of anchoring any validation framework to outcomes seen in the real world.

### *Additional SACATM Comments*

Dr. Clippinger asked if participants in the International Cooperation on Alternative Test Methods[36] will be reviewing the ICCVAM validation document before it is released.

Dr. Page encouraged incorporation of mixture evaluations into new approaches to method validation. Harmonization of categorization will be important to support global use of NAMs. Validation approaches need to consider end use of a substance; what might be considered a non-irritant for a household cleaner may be different from what is considered non-irritant for a face powder. This highlights the need to clearly define the applicability domain of a method. Within the U.S. there needs to be assurance that regional and state agencies are aligned with federal agencies to achieve final acceptance of a method.

Dr. Berg expanded on previous comments on developing biological relevance as well as harmonization. She agreed with the suggestion that specific sets of validation chemicals be established to address specific context of use. These sets of validation chemicals should be diverse enough to address "unknown-unknowns."

Dr. De Abrew wondered whether it might be possible for international collaboration within sectors to help better define biological relevance. Dr. Fitzpatrick noted that a global collaboration is being set up to address this in the food safety area. Dr. Perron added that EPA collaborates with Canada and the European Food Safety Authority, but there can be limitations given the different regulations within each jurisdiction.

## IX. Update on NICEATM Computational Resources

Dr. Helena Hogberg, NIEHS, provided updates on NICEATM's two major computational resources, the Integrated Chemical Environment (ICE)[37] and the Open (Quantitative) Structure–activity/property Relationship App (OPERA)[38].

OPERA provides QSAR-based predictions for a number of chemical properties and endpoints, many of which have been added in the last year. OPERA was built with open-source codes, and its algorithms and performance are very transparent, as are the applicability domains and limitations of the models. It can be run via a command-line interface or a graphical user interface and has a variety of input options. OPERA is now available as a plug-in for the OECD QSAR Toolbox. Predictions from OPERA are available via ICE. Dr. Hogberg reviewed an external publication that used various computational models to predict properties of per- and polyfluoroalkyl substances, where OPERA predictions were found to be more accurate than the ones obtained from other models. A major update of OPERA is in progress.

ICE is a resource for summary-level, high-quality, curated data. It was released five years ago and supports use by a broad target audience, providing interactive and interconnected tools and links to other NIEHS and EPA resources. Dr. Hogberg reviewed the endpoints with data currently in ICE. The most recent update of ICE (v3.7) implemented application programming interfaces (APIs), which support access to ICE

[36] More information at https://ntp.niehs.nih.gov/go/icatm.
[37] Available at https://ice.ntp.niehs.nih.gov/.
[38] More information at https://ntp.niehs.nih.gov/go/opera.

data from outside resources. Recent updates also supported continued implementation of FAIR and TRUST (transparency, responsibility, user focus, sustainability, and technology) standards, new quality control annotations for curated high-throughput screening (cHTS) data, help videos, new and updated data, and a new publications section. NICEATM partnered with PCRM in April to offer training on ICE[39].

Dr. Hogberg then reviewed improvements to specific ICE tools that have been made in the last year.

- Search: improved results graphics.
- Chemical Quest: Saagar fingerprints for structural similarity searching, new filtering options, and new results selection options.
- Chemical Characterization: improvements to product use category characterization and updated documentation.
- Curve Surfer: curve overlay with 3D option and new filtering and selection options.
- PBPK: updated absorbance, distribution, metabolism, and excretion data from httk.
- IVIVE: ability to upload in vivo and in vitro data.

Ongoing work on ICE will incorporate metabolism and population variability into the modeling tools by considering the effects of variations in enzyme metabolism. This activity was motivated by SACATM feedback, as were many other improvements implemented in the last year. Metabolism and population variability will be included in the ICE 3.8 release next spring. Other improvements envisioned for this release include estimation of exposure, improved visualizations of non-cHTS data, updates to models used in the PBPK and IVIVE tools, and updates to OPERA and in vitro dermal data sets.

***Clarifying questions and comments:*** Dr. Gehen asked for clarification on how ADMET Predictor, a proprietary product, will be used in ICE for modeling metabolism and population variability. Dr. Kleinstreuer responded by stating this choice was made after determining that there were no open-source tools that could be used for this purpose. NICEATM is discussing the limits of use with SimulationsPlus, and it is envisioned that ADMET Predictor will most likely be used in ICE on the back end to produce aggregate predictions of metabolism to develop confidence intervals. It's a question of striking a balance between being as transparent as possible and providing useful tools that respect proprietary technology.

## Public Comments

There were no written comments submitted for this section.

### *Oral Public Comments*

Dr. Jessica Ponder, representing PCRM, appreciated the support provided by ICE for the FAIR principles. Rapid development of computational approaches can create a disconnect between the availability of data and the ability to understand them. ICE helps

[39] Materials and videos available at https://pcrm.widencollective.com/portals/ieooh0ol/ICE.

address this disconnect. She encouraged NICEATM in their future development of ICE to address user flexibility in addition to computational complexity. In particular, it would be of interest to update ICE to support users' analyses of propriety data. She stressed that while development of these computational tools is an important advance, there needs to be a continued focus on the goal of communicating the knowledge derived from these analyses to those who make regulatory decisions.

### *Comments from Designated SACATM Discussants*

Discussants for "Computational Resources" were asked to consider the following questions:

- What suggestions for improvements/modifications do you have for an existing ICE interface and/or tools?
- What new functions or tools could be prioritized for future development?
- What other types of data would you like to see in ICE and where might NICEATM obtain them?

Dr. Baran, first discussant, considered these questions in the context of activities that, given limited resources, could best help support improvements in human health. A key theme of the meeting's discussions is how SACATM is in a position to help de-silo efforts and bring people together. He appreciated how the ICE interface has evolved and welcomed implementation of APIs and the increase in training activities. He suggested that ICE could benefit from an increase in computational power, as well as implementation of buttons to provide feedback to the developers in real time about data gaps or usability problems. It would also be helpful to add a glossary of frequently used terminology such as context of use, validation, etc. He cited an example of pharmaceutical industry discussions with FDA around the definition of "microphysiological systems" that helped with engagement and driving progress. Dr. Baran also encouraged development of case studies that explain how tools such as those in ICE advance acceptance of, for example, OECD test guidelines. Additional new functions or tools to consider include checkpoints for discrepancies in data, easily accessible lists of available endpoints, expansion and clarification of metadata, more reference chemical lists, and expanding the scope of data contained in ICE beyond regulatory endpoints. In response to the question on types of data, Dr. Baran stated that there needs to be collaboration across industries to improve access to data, and listed some databases that could have data of interest to ICE. Expanding the data available should be a priority. The gap between the amount of data available and the ability to extract knowledge from that data is huge, and NICEATM can help address that gap. In addition, when considering these data, we need to address qualitative and quantitative variability, and make sure that human diversity is represented. He mentioned an FDA initiative that represented a good example of how to benchmark and standardize data.

Dr. Fourches, second discussant, appreciated the improvements in the user interface in ICE. For OPERA, the standalone version that can be downloaded and used internally is very useful for industry. He noted the improvements to OPERA based on comments from past SACATM meetings.

Dr. Fourches made the following suggestions to improve ICE:

- Add a direct link to the consumer use explorer on the home page.
- Update the search tool to support the use of chemical names, particularly for well-known drugs and pesticides.
- Increase the size of the results windows to help with viewing results.
- Add a filter for "important" results rather than showing all positive and negative results.
- Improve distinctions between predicted and experimental values.

Dr. Fourches also commented that Distributed Structure-Searchable Toxicity substance identifiers are available on the initial results page but not on the detailed results page, making it difficult to access data substances in the EPA CompTox dashboard. He expressed an interest in seeing more structure-based data and information on binding sites and binding modes; linking to the Research Collaboratory for Structural Bioinformatics' Protein Data Bank[40] could help with this. He asked NICEATM to consider how ICE could move beyond just including small molecules but acknowledged that this is a long-term goal. He closed by reiterating his appreciation for how responsive NICEATM has been in incorporating SACATM suggestions for improvement of ICE and OPERA.

### *Additional SACATM Comments*

Dr. Page encouraged NICEATM to provide training on ICE to regulators and industry. She agreed with Dr. Baran about collaborating across sectors to expand the applicability domain of ICE tools, though she questioned the usefulness of incorporating chemical name identifiers. A long-term goal worth addressing is using computational tools to model mixtures. Dr. Fourches replied that supporting chemical name searches might support use by students and educators, and that it might be possible to incorporate broadly accepted chemical names. Dr. Kleinstreuer noted that supporting chemical name searches is something that NICEATM has considered in the past, and that there may be an opportunity to leverage existing chemical synonym data in the EPA CompTox dashboard.

Dr. Berg agreed that improvements could be made to make ICE more accessible to broader audiences by getting input from broad audiences on updates. She also suggested considering ways to make metadata more visible and accessible. She asked what information NICEATM has on ICE use and suggested that broader publicity of ICE is needed. Dr. Kleinstreuer agreed that NICEATM could do more cross-sector outreach to publicize ICE's availability.

## X. Adjournment

Dr. De Abrew provided a summary of the meeting and thanked presenters and members for their participation. He invited concluding remarks from SACATM members. Dr. Page noted that this meeting really showcases the progress made by federal agencies in recent years, and again encouraged the establishment of a permanent NICEATM director. Dr. Charest noted that he was rotating off the committee but

[40] Available at https://www.rcsb.org/.

expressed an interest in being involved in future NICEATM and ICCVAM activities.

Dr. Woychik thanked the chair for his leadership and the other departing members for their contributions. Dr. Brownlow thanked all participants and support staff. Slides from the meeting will be made available on the NTP website when they meet government accessibility guidelines, and attendees will be notified when slides and minutes are available.

Dr. De Abrew adjourned the meeting at 3:35 p.m.

K. Nadira De Abrew, PhD

SACATM Chair

Date: 6 February 2023