STATE OF THE SCIENCE ON DEVELOPMENT AND USE OF NEW APPROACH METHODS (NAMs) FOR CHEMICAL SAFETY TESTING

Conference Summary
December 17, 2019
U.S. Environmental Protection Agency
Washington, DC

Welcome and Charge to the Group

Rusty Thomas (Director of EPA's Center for Computational Toxicology and Exposure [EPA-CCTE]) opened the workshop, welcomed everyone, and introduced Alexandra Dunn (Assistant Administrator of EPA's Office of Chemical Safety and Pollution Prevention). A. Dunn noted that this would be an annual event as EPA aims to be a leader in the world of New Approach Methods (NAMs) and wants to increase clarity and transparency in their application and use. The participation of people from across the Agency and world shows how timely, relevant, and important the topic is. A. Dunn described how Andrew Wheeler (EPA Administrator) has challenged the agency to accomplish the broad and ambitious goals articulated in the September 2019 memo, including reducing its requests for, and funding of, mammal studies by 30% by 2025 and eliminating all mammal study requests and funding by 2035. A. Wheeler's leadership and vision will ensure the success of this cause. A. Dunn then introduced A. Wheeler.

A. Wheeler stated that the charge of the workshop is of personal interest to him, and the conference will bring together all stakeholders to have the conversation and drive the goal forward. EPA has set aside $4.25 million for grants to universities to begin research on NAMs, and A. Wheeler has directed EPA's Office of Pollution Prevention and Toxics (OPPT) to work towards demonstrating measurable impacts on animal testing while continuing to protect human health. Over the past several years, EPA has already made progress in creating NAMs and reducing animal testing across the agency and has done this with the support of external stakeholders. The discussions of the conference are expected to lay the foundation for EPA to continue to pursue the use of NAMs as a replacement for mammal testing, and while the challenges will not have easy solutions, EPA has the power and talent to find the best way to use NAMs. A. Wheeler noted that EPA can and will eliminate mammal testing while maintaining the scientific standards that EPA is known for.

Establishing Baselines for Animal Use at EPA and Opportunities for Reduction

After A. Wheeler's remarks, R. Thomas noted that if EPA is to achieve the goals laid out by the Administrator, EPA needs to know where the field currently stands. The first session is intended to establish baselines and opportunities to move forward.

Anna Lowit (EPA): Retrospective analysis of the statutory requirements, study requests, and research utilization in OCSPP and ORD

Anna Lowit (Science Advisor for EPA's Office of Pesticide Programs) presented a broad overview of "where we are and what we can do" from the perspective of EPA.

Summary

- The various offices across EPA require, request, and use animal tests for different reasons. Some parts of the agency are doing well in terms of already being on track to meet the Administrator's goals, and other parts of the agency still have a ways to go.
- In ORD, use of mammals has already decreased by 50% over the last three years, and there is reason to believe that ORD can decrease by another 30% by 2025.
- In OPP, the number of pesticide submissions varies, but the total number of vertebrates used each year in submissions typically ranges from 20,000 over 100,000.
- Using HASPOC Waivers, over 200,000 animals have been saved by OPP over the past 6.5 yrs.
- In OPPT, the Strategic Plan to Promote the Development and Implementation of Alternative Test Methods

was released which identifies current/ near-term (<3 years) and intermediate-term (3-5 years) activities
- OPPT is near completion of a retrospective analysis of TSCA available, requested, and required information (ATARI)
- There are ongoing cross-cutting collaborations that will inform the use of NAMs moving forward. Multi-stakeholder collaboration will be essential moving forward.

Additional points made:
- The Administrator has charged EPA to develop an action plan in 2020 to reach the two goals; this plan will include elements such as validation of NAMs.
- None of these projects are being run without input from other groups, including across EPA program offices, states, industry, animal welfare organizations, etc.
- The Office of Research and Development (ORD) is at the cutting edge of developing tools like ToxCast, the CompTox Dashboard, and read-across approaches among others.
- Pesticide registrations require various types of animal testing, and FIFRA facilitates the ability to issue data call ins. The Administrator limited the goal to mammals, but some areas are limiting birds and fish as well, and testing waivers have been granted in the past.
- There has been a pivot towards high-throughput assays and computational models to reduce animal testing in the endocrine disruptor screening program, though only 50 or so chemicals have been evaluated so far.
- OPP is already actively engaged with stakeholders to conduct retrospective analyses and to develop or implement NAMs across various types of toxicities.

Questions following the presentation included:
- **George Daston (Procter & Gamble):** I want to commend EPA for its leadership in this area, and much work would not have been done without EPA. I wanted to say that this is a bold initiative and you even used the word "moonshot." With that in mind, the $4.25 million is nice, but European Chemicals Agency has already spent $90 million for similar efforts. Is there a commitment to increase the funding over time so that we have a chance of meeting the 2035 goal?
    - **Jennifer Orme-Zavaleta (EPA's Office of Research and Development [ORD])**: The $4.25 million was just a startup grant fund. There are additional resources and projects that are already being invested in through other NAMs areas, in our new Center for Computational Toxicology and Exposure, and other partnerships (e.g., Health Canada). As the appropriations budgets come in over the next few weeks, hopefully a larger budget will come in for EPA. All of this will help build the road map but it will take a village to achieve the goal. The partnerships and proof of concepts will be the way to drive this effort forward.

Variability and Relevance of Current Animal Tests and Expectations for NAMs

Thomas Monticello (Amgen): Concordance of the toxicity of pharmaceuticals in animals and human: The IQ DruSafe Translational Database

Thomas Monticello (Amgen) presented on behalf of the International Consortium for Innovation and Quality in Pharmaceutical Development (IQ) Consortium, which is composed of over 35 pharmaceutical and biotechnology companies whose mission is to advance science and technology to develop transformational solutions that benefit patients, regulators and the broader research and development community.

Summary
- Current non-clinical testing paradigm assumes animal models and study designs are predictive of possible human hazards, but few publications exist to address concordance between observed toxicities in animals to those in the clinic.
- IQ Consortium developed prospective, blinded database of 182 molecules with animal toxicology together with human Phase I clinical data.
- Placed most emphasis on positive and negative predictive value since they are more aligned with non-clinical to clinical translation.
- Clinical prevalence has a big impact on positive predictive value.

- Positive predictive value for combined endpoints ranged from <30% for rodents to ~45% for non-human primates
- Negative predictive value for combined endpoints ranged from 85-90% for all species. Negative predictive value increased with 'other' category of effects (e.g., headache, fatigue) was eliminated.
- Current animal models do a better job at predicting 'safety' (i.e., the absence of an effect) than specific hazards.

Additional points made:

- The DruSafe database includes prospective clinical and nonclinical data for almost 200 compounds, to evaluate the concordance between nonclinical animal testing results and clinical testing outcomes.
- The investigators found that clinical prevalence rate is the most important variable in the predictive value of animal testing, and the lower the prevalence the lower the concordance. Animal study designs are predisposed to identifying false positives due to the design of the studies, and they do not predict the more subjective outcomes (e.g., headaches). If animals are dosed at high levels and no adverse effects are seen, that is a good predictor of a lack of adverse events in clinical settings.
- Specifically, the absence of animal toxicity equates to absence of toxicity in the clinic, strongly supporting the value of animal testing to provide safety in Phase I drug trials.

Questions following the presentation included:

- **Kathryn Page (The Clorox Company):** You have clinical data to support your in vitro and alternative approaches. How do you think we can use data from the pharmaceutical industry, which has human data, to benefit us in the environmental space?
    - **Thomas Monticello (Amgen):** I think it's going to be difficult. In clinical settings you can start with adverse events from humans and work backwards. With the environmental space you will be trying to screen chemicals out prior to knowing the clinical outcome in humans, which is a challenge in developing and using NAMs. And even in the pharmaceutical area we still have to decide if we are going to use animal cells or human cells, because even extrapolating from animal to human doesn't work that great.
- **Ivan Rusyn (Texas A&M University [TAMU]):** I appreciate IQ taking on this particular dataset because it goes beyond pharmaceuticals as well. But I wanted to point out that for environmental chemicals and industrial chemicals, the database of at least cancer concordance is available, as it was just published last month by D. Krewski as first author. It was based on the re-evaluation of IARC data and showed that every compound that has been found to be carcinogenic in humans had a positive animal study. So I think we need to look more into the negative predictive value, where if a chemical is clean in animal studies, then we can trust the data, but if there is a signal, it really needs to be investigated so we can try to understand what the signal is.
- **George Daston (P&G):** I was really pleased to see the negative predictive value numbers because I think it is going to be important in how we apply nonanimal methods. An observation is that in the case of pharmaceuticals, 100% of the compounds that are being advanced to the clinic were developed for a clear endogenous molecular target. That may not be the case for some unknown fraction, probably the majority, of the commercial commodity chemicals that are regulated by TSCA. So having the confidence that a negative conclusion is also going to be negative in the human population becomes very powerful in terms of how you design and interpret the results from an alternative testing strategy.
- **Rusty Thomas (EPA-ORD):** Would you expect differences in the negative predictive value between the more selective molecules in the drug space compared to the "dirty" molecules in the commodity chemical space? Right now our risk assessment paradigms manage for that negative predictive value by identifying a safe dose and managing toward the safety of the chemical.
    - **Thomas Monticello (Amgen):** I wouldn't think there's any difference. For pharmaceuticals, you know the target and you have a biomarker to see how much that target is engaged, and if it is over-engaged that results in exaggerated pharmacology. So we know which target-related events are going to happen and we can look for off-target toxicity. I would say it is the same with your space, where you are still looking for off-target activity – toxicity or adverse outcomes.
    - **Rusty Thomas (EPA-ORD):** But how does negative predictive value work in a dose context?
    - **Thomas Monticello (Amgen):** We look at the area under the curve (AUC), and look at exposure levels in the animals where you saw the toxicities and the highest dose achieved in phase 1 trials, and

false positives occurred when animals are dosed much higher than AUCs achieved in the clinic.

Nicole Kleinstreuer (NICEATM): Variability of animal studies for acute toxicity, skin sensitization, and mechanistic responses

Nicole Kleinstreuer (NICEATM) discussed reproducibility and variability in animal tests.

Summary

- The qualitative reproducibility of animal hazard data is generally between 70-80% (e.g., Hershberger, 72%; Skin Sensitization, 78%; Uterotrophic, 74%).
- For potency categorization in eye irritation, the reproducibility is dependent on the category – higher reproducibility for the highest potency (Type 1) and non-irritants while mid-potency categories (Type 2A and 2B) are poorly reproducible.
- In acute oral toxicity tests, the 95% confidence interval for LD50 values are generally $\pm$ 0.3 $\log_{10}$ units with highest variability in the high EPA hazard categories.
- For skin sensitization, NAMs perform as good as or better than animal models – hazard (74% vs 80%), 3 class potency (59% vs 60%).

Additional points made:

- For binary hazard endpoints, reproducibility is 70-80%, in particular within the acute space. For potency, the entire range of potential effects is possible when studies are redone. It is therefore important to consider how to identify reproducible data to validate NAMs.
- Some example approaches include identifying chemicals with reproduceable effects in the literature and quantifying confidence ranges across repeated tests. For chemicals that varied across tests, features such as structure, phys-chem properties, and bioactivity were evaluated, but did not appear to drive variability.
- Skin sensitization is a good example of benchmarking alternative methods, where reproducibility and predictivity were similar between the NAM and animal test and may serve as an example or proof-of-concept to show how to benchmark NAMs.
- With respect to developmental toxicity and cancer, there are efforts underway to map mechanisms to human toxicity and establish scientific confidence in these ToxCast and other high-throughput methods.

Questions following the presentation included:

- **Ivan Rusyn (TAMU):** What about strain or other types of variability that come into play not only with setting the safe exposure limit but also with identifying hazard? A number of studies have shown that sometimes particular strains may not have been good models for the disease state of interest, but if you look deeper you can find other strains that are good models for the disease or toxicity of interest. Unfortunately that would require more animals or more complex human models that take into account population variability on the hazard side.
    - **Nicole Kleinstreuer (NICEATM):** Yes, that is a good point. Most of what I showed was based on existing regulatory guidelines which are confined to particular strains and study protocols. But there has been a lot of work that has shown the importance of looking at variability across a heterogeneous population. I would argue there is an ability to look at this in cell-based models, such as by using cell lines that are derived from different strains. And then you can use modeling to estimate confidence intervals that you think adequately represent the variability that you expect to see across the population.
- **Mel Andersen (ScitoVation LLC):** Since you brought up adverse outcome pathways (AOPs) at the end, my question is are NAMs intended to predict toxicity or are they intended to ensure safety? I think we have an opportunity to not do what we have done before of predicting toxicity and what will happen at high doses, when the one-to-one comparisons like you have described are essential. They may not be essential if we start to ask about preventing the key element of a pathway that will prevent downstream outcomes and targeting safety.

Katie Paul Friedman (EPA): Qualitative and quantitative variability of repeat dose animal toxicity studies

Katie Paul Friedman (EPA) discussed the implications of quantitative variability in repeat dose toxicity studies for scientific confidence in NAMs.

Summary

- EPA ORD developed a curated database of legacy animal toxicity studies (ToxRefDB) that currently contains >5000 studies on >1000 substances.
- Applied multiple statistical approaches to evaluate the quantitative variability in repeat dose animal studies.
- Maximal R-squared for a NAM-based predictive model of systemic effect levels may be 55 to 73%; i.e., as much as 1/3 of the variance in these data may not be explainable using study descriptors.
- The estimate of variance (RMSE) in curated LELs and/or LOAELs approaches a 0.5 log10-mg/kg/day.
- Estimated minimum prediction intervals for systemic effect levels are likely 58 to 284-fold based on RMSE estimates.
- The current LOAEL-NOAEL uncertainty factor (UF_L) (i.e., 10-fold) covers the estimated one-sided minimum prediction interval.

Additional points made:

- In silico, in chemico, and in vitro models cannot predict in vivo systemic effect values with greater accuracy than the reproducibility of animal models. To better understand variability in repeat dose studies, what is the range of possible systemic effect values in replicate studies, and what is the accuracy of a model that tries to predict systemic effect values for an unknown chemical?
- ToxRefDB v2.0 can be used to answer some of these questions. Based on the study descriptors in ToxRefDB v2.0, statistical models of the variance in quantitative systemic effect level values could be developed.
- The total variance in systemic toxicity effect values likely approaches 0.75-1 units of $(\log10\text{-mg/kg/day})^2$. The percent explained variance (amount explained by study descriptors) likely approaches 55-73%, which means that the R^2 on some new, predictive model would approach 0.55 to 0.73 as an upper bound on accuracy and as much as 1/3 of the variance in these data may not be explainable using study descriptors.
- Prediction of an animal systemic effect level within -/+ 1 log10 mg/kg/day fold demonstrates a very good NAM, which is important for acceptance of NAMs in chemical safety assessment. Construction of NAM-based effect level estimates that offer an equivalent level of public health protection as effect levels produced by animal-based methods may lead to reduction in the use of animals.

Questions following the presentation included:

- **George Daston (P&G):** How much of the variability is attributed to the limited statistical power of the tests themselves?
    - **Katie Paul Friedman:** It would be great to re-do this with benchmark doses to more fully explore this. Another source of unexplained variance is differences related to methodology, such as different labs or different equipment, but using BMDs or something else that acknowledges that issue could refine the analysis a little bit.
- **Susan Borghoff (ToxStrategies):** When pulling the information together, was there a consideration for characterization or stability of the test substance? In guideline studies that is a requirement, but when you start pulling information from other sources sometimes that information isn't always available.
    - **Katie Paul Friedman:** That would be an issue in some databases; however, most of our replicates tend not to come from the literature space. All of our nonguideline studies were evaluated to determine how guideline-like they were, and one of the questions was whether or not the substance purity was recorded and how the chemical was described in the paper. That is as far as we have gone. We do include substance purity as a covariate in the multilinear regression model so it could explain some of the variance but it is not a statistically significant covariate for us.

State of the Science in Development and Application of NAMs

Dave Allen (ILS): Development of NAMs to predict acute toxicological responses

Dave Allen (ILS) presented a discussion of acute toxicology tests, otherwise known as the "6-pack" (i.e., acute oral,

acute dermal, acute inhalation, skin irritation, eye irritation, and skin sensitization). Between 30-90 animals are used in each 6-pack across all assays.

Summary

- The "six pack" set of acute toxicity study includes acute oral, acute dermal, acute inhalation, skin irritation, eye irritation, and skin sensitization with total animal numbers ranging from 36 – 86 animals.
- Retrospective analysis of data submissions suggested that acute dermal studies provided little additional information enabling the use of waivers.
- Acute oral toxicity data was curated and used to develop consensus QSAR models that performed equally well as *in vivo* studies when variability was considered.
- Acute inhalation toxicity data is being curated and 3D *in vitro* models are being evaluated.
- NAMs for skin and eye irritation are being tested for performance.
- Defined approach for skin sensitization working its way through Organisation for Economic Co-operation and Development (OECD).

Additional points made

- A roadmap was developed by Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) to identify the requirements, needs, and decision contexts for each endpoint, to inform NAM development.
- As low hanging fruit to reduce animal use, certain testing requirements can be waived (e.g., waiving an acute dermal study if an acute oral study has been conducted). There have been workshops that have focused on efforts such as using in silico models to predict acute LD50 values and hazard categories (consensus models were shown to be close to the reproducibility of oral in vivo tests), establishing confidence in in vivo test variability, and ensuring transparency and training across these efforts.
- Data curation is a challenge with developing in silico NAMs, as sometimes databases might be missing data or provide inaccurate data. Efforts to develop an inhalation LC50 model are ongoing.
- In an exploration of eye irritation NAMs, no single test method correctly identified all formulations relative to their in vivo classifications but combining results in an integrated approach may be useful or working towards binary classifications.

No questions followed the presentation due to lack of time.

Tara Barton-Maclaren (Health Canada): Application of NAMs for quantitative screening level risk decisions

Tara Barton-Maclaren (Health Canada) discussed the evolution of risk assessment under Canada's Chemicals Management Plan (CMP), translating case study findings into application, and building confidence in NAMs within the Health Canada framework.

Summary

- The Chemicals Management Plan (CMP) is a program to reduce the risks posed by chemicals to Canadians and their environment. In Phase 3, 1550 priority chemicals out of the original 4300 chemicals will be addressed by 2020. Many of the chemicals do not have traditional animal toxicity data.
- Health Canada releases Science Approach Documents (SciADs) that outline specific scientific approaches which can be used in future assessments or prioritization exercises.
- Developing a SciAD around on a joint EPA-HC-ECHA-ASTAR case study demonstrating that *in vitro* bioactivity from ToxCast provides a conservative estimate of a point-of-departure from traditional animal toxicity studies (chronic, dev, repro).
- Showed that the bioactivity-to-exposure ratios (BER) for a subset of chemicals were generally aligned with CEPA 6(c) assignments and will be used to inform priority compounds under the CMP.
- Developing preliminary uncertainty factors (UFs) to apply to screening level assessments based on *in vitro* bioactivity.
- Working to expand the approach by using bioactivity from nearest neighbors and *in silico* toxicokinetic estimates.

Additional points made:

- A Science Approach Document (SciAD) does not include regulatory decisions but describes novel approaches to evaluate the potential for environmental or human health risks. The use of streamlined

assessment approaches has allowed Health Canada to meet requirements under the CMP; streamlined methods have been applied to about 70% of assessments over the last decade.
- The Risk Assessment Toolbox can help identify the level of effort needed to identify hazard/risk during an assessment.
- Case studies are critical to advancing knowledge, but it is important to translate those findings into applications and practice. One case study is the Accelerating the Pace of Chemical Risk Assessment bioactivity-exposure ratio (BER) retrospective case study, which is a collaborative effort that showed 90% of chemicals had a lower POD based on bioactivity than the traditional Point Of Departure value.
- Health Canada and EPA have already begun characterizing some of the uncertainties with NAMs, including discussions around defining uncertainty factors. However, there are additional data gaps to address to move forward in the application of NAMs, including increasing the chemical space of the BER approach.

Questions following the presentation included:

- **George Daston (P&G):** This is the kind of application that we were hoping NAMs could be used for first in a regulatory context. My question is whether there is anything different or unique about the high-throughput assays that do not seem to behave like the majority, with either a very wide or very narrow BER. Are there specific assays that seem to be highly sensitive or groups of assays that would lead us to suggest certain dose-response characteristics we would want to pay more attention to?
  - **Tara Barton-Maclaren**: I think it's a combination of a number of different things causing these outliers. For example, the chemicals could have been in the more volatile range or had specific chemical properties that impacted the results; the traditional PODs might not have been based on the best data (e.g., the study had outlier dose ranges); or sometimes the exposure data was based on inhalation route but compared to an oral equivalent dose. We are looking at a number of factors more closely to determine what is driving the lower bound. It appears that some chemicals which should be excluded from an analysis like this based on other factors are the ones appearing as outliers.
- **Ivan Rusyn (TAMU):** I think the original impetus for moving from animal studies to something else was that we can't test our way out of the long list of chemicals we need to evaluate by doing animal tests. I get a sense that maybe we can't test our way out using in vitro data either because the field of possible in vitro tests or combinations of in vitro tests is infinity. So do you think we should start gathering up what's been done so far and identifying a limited number of tests that can be applied to a very large number of compounds, or are you still okay with a kind of wild west of in vitro toxicology?
  - **Tara Barton-Maclaren:** I think that question could be answered in many different contexts and depending on your program requirements, so we have to keep that in mind. However, I think there are a number of structural contexts that we have heard about today (e.g., IATA, AOP) which could be the next step following this type of screening and priority setting, to facilitate more targeted testing and addressing the data gaps or the research requirements that we need to target for specific toxicities. In terms of if there's a specific set of tests at this time, I would say no, I don't think we know what that looks like, and I'm seeing heads nod around the room. I think in the case of acute we have seen some promising work but I think there is a lot of work to do as we move into the more chronic endpoints. There is a tremendous amount of data available so we should leverage the information that we know and continue with evaluations to see where we can best focus targeted efforts for those chemicals that may emerge as being of greater concern or for which we know there's direct potential for exposure. We need to continue to consider the risk-based lens to better define what targeted testing requirements could be moving forward.
- **Rick Becker (American Chemistry Council [ACC]):** I think we all benefit from these kinds of case examples. Can we consider the threshold of toxicological concern (TTC) as a kind of starting point? If we can use that with confidence as a kind of screening toxicity value and we have confidence in what we're looking at in terms of the exposure assessment, it seems that we can look at chemicals very quickly and put some in a bin of maybe not needing much more extra work, including performing high throughput or high content testing. Is that correct?
  - **Tara Barton-Maclaren**: That's right. When we looked at the case study we did look at just applying the TTC-based approach as a first screen. And it was interesting to see that the

using the BER approach was a refinement to the results from the TTC. We are certainly looking at developing and using complimentary tools, including QSAR and computational TTC approaches, and another project that we're working on is developing a similar type of workflow to identify those chemicals that may have the potential for genotoxicity by leveraging other tools. It is critical to move forward in a complimentary way so we're covering as much biological and chemical space as possible.

George Daston (Proctor & Gamble): State of the science for predicting developmental toxicity using NAMs

George Daston (P&G) presented on how NAMs can be used for predicting developmental and reproductive toxicity, which is often required in regulatory contexts.

Summary

- Current approaches require a variety of animal tests to cover the entire reproductive and developmental cycle, and current NAMs only cover a part of the cycle.
- The 2017 National Research Council Report suggests combining cheminformatics, pharmacokinetic models, systems biology, and mechanistic models into a predictive toxicology workflow to identify acceptable doses for untested substances.
- The predictive toxicology workflow relies heavily on read across as the primary method for assessing developmental and reproductive toxicants on a broad basis
- A range of *in vitro* assays exist for developmental toxicity including whole embryo culture, stem cell assays, and free living embryos (zebrafish).
- Criteria for believing in a NAM
- Covers a defined range of modes of developmental tox
- Integrated with other assays to comprehensively cover all potential modes of action for dev tox
- Responsive to human developmental toxicants with dose concordance

- To fully evaluate this type of toxicity, it is important to cover the entire reproductive and developmental cycle. Current approaches require a variety of animal tests to cover the entire cycle, and current NAMs only cover a part of the cycle. For new compounds, read across is used prior to conducting additional testing.
- There is a predictive toxicology workflow that can be leveraged for NAM development, which uses cheminformatics, PK models, systems biology, and mechanistic models.
- Structural and biological information can be used to inform read-across decisions. Read across will likely be used moving forward to eliminate animal testing, as it is the best and only method for assessing developmental toxicity on a broad basis.
- NAMs would need to cover a defined range of modes of development toxicity, consistently. They could be combined with other assays that cover the remaining modes of action for universal coverage. And NAMs need to be responsive to human developmental effects.

Questions following the presentation included:

- **Rick Becker (ACC):** What are your thoughts on TTC?
    - **George Daston (P&G):** TTC is the place where you start. I think pragmatically we know from lots and lots of toxicity testing that there is a level below which you really don't see toxicity. There are exceptions to that rule, and we can identify those and exclude those classes of chemicals from TTC, but TTC is a starting place. However, I think for us to fully get beyond animal testing we're going to need more.
- **Mel Andersen (ScitoVation LLC):** Two points come to mind. Going into the future, where would you see the opportunities for looking at mode of action, more generally than just with the reproductive and development end points? And then as we think about reproductive and developmental hazards, what kind of augmentation do we have to do on toxicokinetics in in vitro assays to look at fetal stages of development?
    - **George Daston (P&G):** In terms of mode of action, I think we need to do the hard work of going back to 50 years of traditional toxicology layered on top of all of the data of toxicogenomics and the high-throughput assays to learn what exactly are the biological targets for toxicity. I think it is a different way of looking at the field and rather than organizing our field by endpoint like we do now, we have to organize it by mode of action. In terms of toxicokinetics, I don't think it's going to be that different. I

think that most compounds access the embryonic compartment following the same rules that govern volume of distribution or even absorption, so movement across the placenta is most often going to be governed by things like passive diffusion and then factors like protein binding on the maternal side.

Dan Tagle (National Institutes of Health): Advances in the development of organotypic and tissue chip technologies for toxicity testing

Dan Tagle (NIH) described tissue chip technology which is used by the National Center for Advancing Translational Sciences as an in vitro tool for drug development.

Summary

- Tissue chips are microfluidic cell cultures that build up organ structures from basic functional units such as scaffolding, cells, etc.
- A number of organ platforms exist, from skin, to the gastrointestinal system, nervous system, liver, kidney and more. On a liver chip, for example, the middle of the chip includes hepatocyte cells and non-parenchymal cells to recapitulate aspects of liver function, and a flow culture allows for transport. Toxicity can be measured using imaging and staining.
- Organs on chips are designed to be modular and can be linked together, like a group from Pittsburgh that linked the blood brain barrier, liver chip, and other organs together.
- Initial efforts to develop tissue chips included NIH, FDA and Defense Advanced Research Projects Agency (DARPA). Since then, the partnerships have grown to include other agencies such as NASA.
- Added validation groups to ensure the tissue chips are reproducible and transferable.
- Initiatives at NIH for further development includes representing as much as of the population demographics in the chips as possible, working towards building a human body on a chip, and addressing drug failure rates.

Questions following the presentation included:

- **Unidentified participant:** Where do you think this technology will be in 10 years?
    - **Dan Tagle:** The program has been designed such that we are positioning to make this a type of benchtop research tool, which is part of the reason we're pushing the commercialization aspect. One of the collaborations we have with NASA is to facilitate making the technology much more miniaturized and automated. While the chip is very tiny the supporting instrumentation is quite large in terms of running the system and miniaturizing which would increase ease of use and access to this technology. The technology would help with refinement and reduction in terms of animal use. Some have said that in 10 years this technology should be replacing animals and I don't know if we can live up to that expectation but we're working towards that.

Doug Wolf (Syngenta): Development and application of in vitro methods for evaluating respiratory irritants

Doug Wolf (Syngenta) presented a talk on a fundamental purpose of NAMs, to inform regulatory decision-making in lieu of animal testing.

Summary

- A sub-chronic whole animal inhalation study is a regulatory requirement, but for chlorothalonil, the study was not expected to improve human safety. Therefore, scientists evaluated the possibility of developing a NAM that would be suitable to inform inhalation toxicity.
- Particle size distributions and aerosols evaluated during pesticide applications
- Airway dosimetry estimated using computational fluid dynamic (CFD) modeling.
- A 3D model of the human airway epithelium from 5 different donors exposed for 24 hours at 10 concentrations was developed for this case study.
- Measured trans epithelial electrical resistance, LDH, and resazurin.
- BMD modeling identified a human equivalent concentration.
- The approach addressed the requirements of an inhalation study without killing animals and addressed

uncertainty factors as well.

Additional points made:

- The case study focused on the exposure component and the interaction of the chemical with the respiratory tract and subsequent toxicity. Once the AOP was better understood, potential NAMs could be identified for events along the pathway.
- One NAM evaluated was a 3D in vitro model of the human airway epithelium, which demonstrated good dose-response and consistency in tests, and from which a benchmark dose and then a human equivalent concentration could be calculated. Combining this with exposure information could then inform risk characterization.

There were no questions following this presentation.

Maureen Gwinn (EPA): Identifying endocrine disrupting chemicals using in vitro and computational approaches

Maureen Gwinn (EPA) discussed the Endocrine Disruptor Screening Program (EDSP) and how it has driven the use of NAMs at EPA.

Summary

- The Food Quality Protection Act (FQPA) and Safe Drinking Water Act (SDWA) requires evaluating ~10,000 substances for potential endocrine activity.
- The Endocrine Disruptor Screening Program (EDSP) established a two-tiered system to evaluate chemicals. Tier 1 uses ~600 animals and costs ~$1 million.
- In 2011, EPA began a multiyear transition to prioritize and screen thousands of EDSP chemicals using high-throughput *in vitro* assays and computational modeling approaches.
- Multiple high-throughput screening assays used to screen chemicals for estrogen and androgen receptor pathway activation with computational modeling used to integrate the data. Reference chemicals used to characterize the performance of the computational model/assays.
- Consensus QSAR models developed for ER and AR agonism/antagonism.
- Lessons learned included understanding of the impact of cytotoxicity on assay results, the utility of developing models from a subset of assays, the impact of metabolic competence on assay results, and need to quantify uncertainty.
- Additional activities under EDSP include developing a model for steroidogenesis and thyroid.

- The EDSP program required extensive testing to classify endocrine-disrupting chemicals. EPA developed multiple high-throughput assays to replace animal testing in the battery of required tests, based on a mechanistic understanding of endocrine disruption. EPA evaluated the approach to gain confidence and acceptance for it by comparing assay results to existing peer reviewed publications and submitting the results to scientific advisory panels and OECD review.
- Lessons learned from this effort included a more in-depth understanding of the impact of cytotoxicity on assay results, the utility of developing smaller pathway models from a subset of assays, the impact of a lack of metabolic competence on interpreting assay results, and uncertainty around high-throughput results.
- There are also large-scale QSAR efforts underway to build predictive models for ER and AR activity. These models, Collaborative Estrogen Receptor Activity Prediction Project (CERAPP) and Collaborative Modeling Project for Androgen Receptor Activity (CoMPARA), can make predictions for chemicals not measured in high-throughput assays. Additional activities under EDSP include developing a model for steroidogenesis and using the AOP for the thyroid-outcomes to inform high-throughput screening assay development.
- Next steps include expanding acceptance and implementations of this work through OECD, continuing to apply these approaches to other EDSP needs, translating to tissue- and organ-level effects, and including exposure components to inform the risk context.

Questions following the presentation included:

- **Rick Becker (ACC):** This is a great illustration of the power of NAMs when you have a fit for purpose need and a decision context. In this case we are looking at known modes of action which can make it

easier to target specific biology. My other comment is that I am a proponent of the BER approach and with IVIVE and now with HTIVIVE, we're able to do that more and more. I think we are right on the cusp of being able to merge the bioactivity with the prioritization pretty readily, so I think EDSP will be an important application of the BER approach as we move forward.
    - **Maureen Gwinn**: I think the ER and AR were more straightforward and were the low hanging fruit, although it has been about 10 or 12 years. So I agree that it's a proof of concept of how we can do this, by focusing on AOPs and mode of action.
  - **Tara Barton-Maclaren (Health Canada):** My question is related to the addition of the metabolic competence. I was wondering if a comparison has been done to understand the impact of the metabolic competence, since you mentioned there could be an over or an underestimation without it. I'm curious to see or know if there's a tendency towards one way or the other and what the magnitude of that is.
    - **Maureen Gwinn:** I know we have done a magnitude evaluation I just don't have the numbers. The bottom graphs show that in some cases you can see the parent versus metabolite where you could get there with either the bioactivation or the inactivation.
    - **Tara Barton-Maclaren:** And I asked the question to bring it back to the work that we're doing in terms of being able to screen with that in mind and then also be able to identify the classes of chemicals which are more likely to be underestimated and apply uncertainty factors or those kind of caution flags appropriately for that broader application.

Developing Scientific Confidence in NAMs

Warren Casey (NICEATM): New Approaches to Validation and Characterizing Performance of NAMs

Warren Casey (NICEATM) presented on validation of NAMs.

Summary

- The traditional model of validation based on OECD GD 34, which is the most widely used model available, does not work. Issues include segregation of steps, which causes a lot of time and effort to be wasted.
- When thinking about validation, it is important not to think linearly, but to think about guiding principles.
- It is important to start with the end use in mind for validation. That will inform the agency or industry needs
- Common roadblocks to uptake of NAMs include institutional resistance, using animal data as the reference for validation, and harmonization.
- Necessary changes in the validation process include transitioning from centralized to more individual users, moving away from one-size-fits-all towards fit for purpose validation, moving away from seeing something as validated or not and towards increasing confidence over time, and moving away from standalone validation towards more integrative methods.

Additional points made:

- For new approaches to validation, EPA is ahead of the curve. For the acute 6 pack studies, the majority of animals will be saved by looking at what information is actually needed, and data waivers and QSAR models can be used. There is a constant information exchange to inform the validation process.
- Part of the validation process needs to include methods on how the wider audience will use the methods.

Questions following the presentation included:

- **Doug Wolf (Syngenta):** What do you see as the process for addressing or pushing back on all of the methods that people are creating, for example in the realm of funding agencies or organizations or at the university level, so they understand the rules of the road with respect to validation and are not developing new methods without understanding the why?
  - **Warren Casey:** It starts with identifying the actual needs of how the data will be used and what the regulators actually need, not just which tests are done. The end users both in industry and within the regulatory agencies need to clearly explain their needs. We're in the process of developing webinars to help bring in regulators and industry stakeholders to communicate with the people that that are developing these methods and the funding agencies on what is actually needed.

- **George Daston (P&G):** One of the impediments to moving forward is we try to replace the assays that were developed in the 1960s with other assays, and I think it is time to take a step backward and say we were trying to protect public health by identifying agents that interact with biological systems, and the read out then happened to be adverse outcomes. The read outs now are very different, and I think we face a problem in trying to explain to people who become accustomed to looking at adverse outcomes that they can get the same level of public health protection by looking at biological activity. How many regulators do you think are ready to hear that message?
    - **Warren Casey:** I know of one, which is what makes it truly remarkable. We are bending over backwards trying to figure out how to make this work. The only way you have the option of trying something different is by having regulators that dedicated and involved.
- **Rick Becker (ACC):** Your point about starting with the end in mind sounds simple but it is very challenging to catalyze change. It is helpful to have the roadmap. The approach in OECD has helped with mutual acceptance of data globally, as a standardized test guideline created this treaty of mutual acceptance of data. I think we need to start talking about these things and build out case examples and it will help to increase mutual acceptance of data like what is expected from a standardized test guideline.
    - **Warren Casey:** There are also other things within OECD which can slow things down, but historically this has always been a problem.
- **Rusty Thomas (EPA-ORD):** Related to having more of a continuous evolution of confidence – I think the difference between countries can be a problem. Confidence and comfort in acceptance of data can mean different or the same things across countries, and scientific confidence could be a regulatory justification in one country and not in the other. There is very little harmonization with levels of confidence in these things. How would you see harmonization across these different areas and countries?
    - **Warren Casey:** EPA needs to continue taking leadership positions aligned with Canada which is very forward-thinking, whether or not there is an OECD test guideline acceptance. Identify a coalition of the willing to represent a significant coalition of OECD countries, and that might be a way to move things forward. A number of different models start with EPA taking leadership and a coalition of people who want to collaborate with EPA.
- **Mel Andersen (ScitoVation LLC):** When I think of the problem of convincing people such as assay developers and regulators that something is valid, there is still a large portion of the public who think that we are not doing a good job of not protecting them. As you develop approaches to bring stakeholders to the table, how are you doing the outreach to the public?
    - **Warren Casey**: Part of this whole process is acknowledging risk communication to the public, which is something that we need to work on. New processes come with new risk communication plans, and they have put a lot of effort into it.
    - **Anna Lowit (EPA-ORD):** NGO groups came out saying things similar to this that the EPA is not protecting the public with these efforts, but it is categorically incorrect that EPA is not moving methodically. EPA is using FACAs, advisory panels, stakeholder groups, nonscientific groups, and others to inform this work. There are presentations and webinars to get the word out. Any change is hard but it makes the science better, and then we can make better public health decisions. Making sure to put out the best science, using diverse points of view and getting it to as many people as possible, is how we move forward.