Environmental Protection Agency - Pesticide Program Dialogue Committee
21st Century Toxicology/New Integrated Testing Strategies

Where Vision Meets Action: Practical Application of 21st Century Methods
July 9th, 2013 Stakeholder Workshop
Potomac Yard South, First floor conference room
The Workshop will be accessible by webinar at:
https://epa.connectsolutions.com/tox21century/
Call in number: 866-299-3188, code 703-308-0293

Jennifer Sass, Ph.D.
Senior Scientist, Natural Resources Defense Council (NRDC) and,
Professorial Lecturer, George Washington University

There are a number of initiatives across federal agencies to advance predictive toxicology and exposure science:[1]

- EPA's computational toxicology (CompTox) program is investing in computerized models and programs to identify toxicity pathways and thereby increase the ability to identify upstream predictors of adverse outcomes.
- The ToxCast program is focusing on cell and non-cell assays to identify toxicity pathways and develop "toxicity signatures" for specific health outcomes or endpoints. It is screening about 300 data-rich chemicals and about 750 less-characterized chemicals through the same 600 high throughput assays.
- ExpoCast is an EPA program invested in developing high throughput exposure models. It is using CDC NHANES biomonitoring data to ground-truth the models.
- Tox21 is an inter-agency program screening 10,000 chemicals with 50 assays.

EPA notes that its integration of new data into risk assessment or regulatory decisions will be step-wise and iterative, and initially it will be used for prioritization of chemicals for the TSCA inventory or EDSP (endocrine disruptor screening program). [2] Other possible uses may be in the review of pre-manufacture notices (PMNs) for new chemicals registered under TSCA, and to help sort chemicals into categories based on similar biological activity.

[1] McPartland J. Chemical testing in the 21st Century: Opportunities and challenges. Meeting report. Environmental Defense Fund. January, 2013. http://www.edf.org/health/chemical-testing-21st-century
[2] McPartland J. Chemical testing in the 21st Century: Opportunities and challenges. Meeting report. Environmental Defense Fund. January, 2013. http://www.edf.org/health/chemical-testing-21st-century

Benefits of and Challenges to Implement Alternative 21st Century Methods

There are two main recognized benefits to the implementation of data from non-whole animal or alternative testing methods (cell-based testing, cheminformatics, computational toxicology, etc): first, rapid and less-expensive testing methods, and second, methods that can provide information about potential mechanisms or molecular pathways of toxicity. These two characteristics lead to all sorts of advantages. For example, because tests may be faster and less costly, it is possible to test a greater number of chemicals, chemical mixtures, more chemical concentrations, and chemicals in various carrier substances. Because tests can indicate pathways of toxicity, then may be used to predict downstream adverse outcomes, chemical interactions such as additivity, or the potential impact of genetic variability on toxicity.

However, the ability to benefit from alternative testing methods comes with some significant challenges. Some of these include the lack of metabolic activity in the test system (which could over- or under-predict the toxicity of chemicals), the inability to assess latent disease outcomes or the effects of long-term exposures, and the poor ability to assess systemic toxic effects such as adverse impacts on learning, behavior, or memory, birth defects, reproductive impairments, or second generation effects.

Before EPA can rely on alternative tests they will need to be validated. All the experts recognize that validation of alternative methods is a serious issue – possibly the most serious. If EPA does not validate the tests against reliable information, then they are not valid. What constitutes reliable information? Alternative methods should be validated against data from well-designed and well-conducted studies – usually whole animal or epidemiologic studies – with adequate statistical power, covering an adequate range of human exposures, and that exclude with reasonable certainty bias, confounding, and chance. The framework for Systematic Review and Evidence Integration being developed by the NIEHS Office of Health Assessment and Translation (OHAT) provides a powerful approach for evaluating studies, including non-animal or alternative studies.[3]

It is important to recognize the distinction between evidence that something is harmful, inadequate evidence to make a determination, and evidence suggesting lack of toxicity, or "safety". The criteria for the latter – proving a chemical is "safe" – are intentionally stringent for agencies tasked with the responsibility of protecting health and the environment (see for example IARC Monograph classifications, [4] or draft OHAT framework for systematic review[5]). Serious public health consequences may follow if chemicals are misclassified as less toxic or non-toxic based on untested mechanistic hypotheses, poorly validated tests, or incomplete

[3] http://ntp.niehs.nih.gov/?objectid=960B6F03-A712-90CB-8856221E90EDA46E
[4] http://monographs.iarc.fr/ENG/Classification/index.php
[5] http://ntp.niehs.nih.gov/?objectid=960B6F03-A712-90CB-8856221E90EDA46E

data sets. [6] Declaring a chemical as not hazardous, or reducing a level of health protection, should require validation, not speculation.[7]

Building Confidence in the Regulatory Application of Alternative 21st Century Methods for Evaluating Pesticides

The integration of data from non-whole animal or alternative testing methods will be closely scrutinized by an interested public. We suggest that EPA not set out to re-invent the wheel, but rather adopt the frameworks that are now being developed. In particular, NRDC strongly support the framework for Systematic Review and Evidence Integration being developed by the NIEHS. OHAT provides an elegant framework – simple, functional, and effective - for enhancing transparency and communication, promoting consistency, and facilitating reproducibility across literature-based evaluations of hazardous chemicals. It is being developed with transparency and opportunities for public input. NRDC provided comments on the draft OHAT framework (June 11, 2013).[8]

Establish standard information requirements to evaluate study findings and risk of bias for in vitro studies

The OHAT framework wisely sets different criteria for evaluating data from different sources. For an in vitro study the framework assesses the model (species) and treatment, the endpoint of interest, and the concentrations tested. For model and treatment, the following information is collected: [9]

Species;
Cell-line/Source; Sex;
Concentrations tested;
Purity and Source of test agent;
Vehicle for test agent;
Treatment Period;
Replicates; Funding Source;
Author conflict of interest;
Comments (shape of dose-response curve, etc).

This information is collected for each study, and presented in a tabular format. EPA should require this information at a minimum.

EPA should evaluate "believability" of study results, rather than internal validity, reporting quality, GLP compliance, or ToxRTool

[6] Melnick RL, Kamel F, Huff J. Declaring chemicals "not carcinogenic to humans" requires validation, not speculation. Environ Health Perspect. 2003 Apr;111(4):A203-4.
[7] Melnick RL, Kamel F, Huff J. Declaring chemicals "not carcinogenic to humans" requires validation, not speculation. Environ Health Perspect. 2003 Apr;111(4):A203-4.
[8] http://ntp.niehs.nih.gov/?objectid=960B6F03-A712-90CB-8856221E90EDA46E
[9] See draft BPA protocol (April 9, 2013), Table 8, page 34.

For assessing the quality of individual studies, EPA should consider the risk of outcome-specific bias using five domains: selection bias, performance bias, attrition bias, detection bias, and reporting bias. We support this as reflecting the state of the science for approaches to identifying bias in a systematic and transparent manner. EPA should consider more than reporting quality for a measure of study quality for animal and human studies. Reporting quality is not the same as internal validity, which is a better measure of study quality.

If EPA were to only rely on reporting quality as a measure of study quality, it would favor/bias towards GLP-compliant (Good Laboratory Practice) studies, when, ironically, the GLP-compliant studies may actually be the most likely to be insensitive to health endpoints being measured. "Good Laboratory Practices" is a standard for animal care and data collection required for industry laboratories in response to fraudulent practices documented in the 1970s. Industry-funded studies are required by EPA and FDA to follow so-called Good Laboratory Practices (GLP) standards, which include specified approaches to recordkeeping to facilitate audits and reduce fraud (54 Fed. Reg. 34034 (August 17, 1989). GLP requirements are not necessarily associated with higher quality research, proper study design or correct statistical analysis.[10] In most cases, GLP studies have not even undergone scientific peer-review and publication. GLP studies are most often designed to identify major toxic effects (apical effects) like cancer. The problem is that major (apical) endpoints will not be predictive or indicate early-warnings of potential toxicity leading to "major" adverse health outcomes. GLP studies don't necessarily use modern methods for evaluating chemicals and aren't designed to grapple with the problems of low-dose exposures, endocrine or hormonal effects, behavioral or learning effects, immunotoxicity, cardiotoxicity, or upstream effects like reduced sperm count or reduced anogenital distance which are predictors of infertility.

The ToxRTool was developed to assess the reporting quality of a study, and is not an appropriate measure of the internal validity of *in vitro* studies or of risk of bias or overall study quality. The reliability categories utilized in the ToxRTool are the same as the Klimisch codes of reliability (Klimisch et al. 1997). Since the Klimisch codes favor GLP compliant studies, than using ToxRTool would be subject to the same criticism as using either Klimisch codes or GLP as measures of study quality.

Mechanistic data could raise the hazard identification,
but should not be used to explain away hazard evidence.

EPA should consider – but, not requiring – mechanistic data where available as part of the overall evaluation, but not as any more or less valuable than other evidentiary data, including but not limited to important evidence in populations from occupational or environmental epidemiologic studies. Mechanistic (or mode of action, MOA) data should be treated as a parallel stream of data, and can be very helpful in interpreting human and animal data. However, mechanistic data should not be seen as either necessary or sufficient for interpreting

[10] Myers, J. P., F. S. vom Saal, et al. (2009). "Why public health agencies cannot depend on good laboratory practices as a criterion for selecting data: the case of bisphenol A." Environ Health Perspect 117 (3): 309-15.

or evaluating other data. Further, EPA should dismiss arguments about MOA that are really arguments about the potency or degree of risk, and are therefore not relevant to a simple hazard determination.

Exclude underpowered studies that fail to find an effect (null-association), but not studies that find an effect despite being underpowered

For a continuous endpoint, EPA should determine if the study was adequately powered, and if the results are consistent across studies. If there is inconsistency across studies, and it's the underpowered studies are showing null association, than they should be excluded from consideration. In other words, EPA should eliminate null-association studies that are inconsistent and underpowered, but not eliminate studies that may be underpowered if they do find and effect. This is because an underpowered study that fails to find an effect cannot be interpreted, but an underpowered study that does find an effect indicates that the effect is real. As an analogy, if you reach into a haystack a few times (an underpowered study) and don't find a needle (a null study), you cannot conclude whether or not there may be needles in the haystack, whereas if you do find a needle (an underpowered study that finds an effect), than there is at least one needle, and probably more, in the haystack (the effect is real). This is particularly relevant to evaluating epidemiologic studies, where confidence in exposure assessments is limited by the range and duration of exposure in the studied population.

When information is missing or unreliable, EPA should use established defaults that will protect health, and set stringent criteria for when to depart from health-protective defaults

The use of *in vitro* and CompTox mechanistic/MOA data must be interpreted relative to the plausibility of the established default, and not as if the alternative to the proposed mechanism/MOA were no proposed mechanism/MOA. When information is missing or unreliable, EPA should be clear and consistent that its approach is to use scientifically-based default assumptions that will protect health to improve the timeliness of the chemical assessment and decision-making process, and set clear scientifically-based criteria for when to depart from these assumptions.[11]

In the landmark "Science and Decisions" report (NAS, 2009), the NAS committee concluded that, "established defaults need to be maintained for the steps in the risk assessment that require inferences."[12] The NAS committee recommended that EPA and other agencies update default factors and assumptions based on the best current science, identify where unstated or implicit assumptions are used, and replace these with explicit assumptions wherever possible. These recommendations push Agencies to, "continue and expand use of the best, most current

[11] NRDC Issue paper. Strengthening toxic chemical risk assessments to protect human health. S Janssen, J Sass, T Schettler, G Solomon. February, 2012.
http://switchboard.nrdc.org/blogs/jsass/nrdc_issue_paper_better_risk_a.html

[12] Science and Decisions: Advancing Risk Assessment. National Research Council of the National Academies. (2009), p. 7.

science to support or revise its default assumptions,"[13] making the assumptions stronger, rather than reducing reliance on them. In fact, the committee specifically recommended that EPA develop "clear standards for departures from defaults."[14] The committee also noted that establishing, "clear criteria for departure from defaults can provide incentives for third parties to produce research" that can reduce uncertainty and, over time, result in more accurate assessments.

Importantly, by using the established defaults more often, EPA could avoid "the delay entailed by having to re-examine generic information with every new risk assessment."[15] EPA should also evaluate and quantify, when possible, the impact of the uncertainty associated with a default assumption, including a description of how using a default versus the chosen alternative assumption affects the decisions that protect the environment and public health.

[13] Science and Decisions, p. 207.
[14] Science and Decisions, p. 199.
[15] Science and Decisions, p. 191.