



Workshop Report

Evidence-based Toxicology for the 21st Century: Opportunities and Challenges*

Martin L. Stephens¹, Melvin Andersen², Richard A. Becker³, Kellyn Betts⁴, Kim Boekelheide⁵, Ed Carney⁶, Robert Chapin⁷, Dennis Devlin⁸, Suzanne Fitzpatrick⁹, John R. Fowle III¹⁰, Patricia Harlow¹¹, Thomas Hartung¹, Sebastian Hoffmann¹², Michael Holsapple¹³, Abigail Jacobs¹¹, Richard Judson¹⁴, Olga Naidenko¹⁵, Tim Pastoor¹⁶, Grace Patlewicz¹⁷, Andrew Rowan¹⁸, Roberta Scherer¹, Rashid Shaikh¹⁹, Ted Simon²⁰, Douglas Wolf¹⁴, and Joanne Zurlo¹

¹Johns Hopkins University, Baltimore, MD, USA; ²Hamner Institutes for Health Sciences, Research Triangle Park, NC, USA;

³American Chemistry Council, Washington, DC, USA; ⁴Freelance science writer and editor, Takoma Park, MD, USA;

⁵Brown University, Providence, RI, USA; ⁶Dow Chemical, Midland, MI, USA; ⁷Pfizer Global R&D, Groton, CT, USA;

⁸ExxonMobil, Irving, TX, USA; ⁹Food and Drug Administration, College Park, MD, USA; ¹⁰Environmental Protection

Agency, retired, Pittsboro, NC, USA; ¹¹Food and Drug Administration, Silver Spring, MD, USA; ¹²seh consulting + services,

Paderborn, Germany; ¹³Battelle, Columbus, OH, USA; ¹⁴Environmental Protection Agency, Research Triangle Park, NC, USA;

¹⁵Environmental Working Group, Washington, DC, USA; ¹⁶Syngenta, Greensboro, NC, USA; ¹⁷DuPont Haskell Global

Centers for Health and Environmental Sciences, Newark, DE, USA; ¹⁸Humane Society International, Gaithersburg, MD, USA;

¹⁹Health Effects Institute, Boston, MA, USA; ²⁰Ted Simon, LLC, Winston, GA, USA

Summary

The Evidence-based Toxicology Collaboration (EBTC) was established recently to translate evidence-based approaches from medicine and health care to toxicology in an organized and sustained effort. The EBTC held a workshop on “Evidence-based Toxicology for the 21st Century: Opportunities and Challenges” in Research Triangle Park, North Carolina, USA on January 24-25, 2012. The presentations largely reflected two EBTC priorities: to apply evidence-based methods to assessing the performance of emerging pathway-based testing methods consistent with the 2007 National Research Council report on “Toxicity Testing in the 21st Century” as well as to adopt a governance structure and work processes to move that effort forward. The workshop served to clarify evidence-based approaches and to provide food for thought on substantive and administrative activities for the EBTC. Priority activities include conducting pilot studies to demonstrate the value of evidence-based approaches to toxicology, as well as conducting educational outreach on these approaches.

Keywords: evidence-based toxicology, evidence-based medicine, systematic review, validation, test method assessment

1 Introduction

Formed in 2011, the Evidence-based Toxicology Collaboration (EBTC) is a consortium of stakeholders from academia, government, and industry promoting the use of evidence-based approaches in toxicology. The EBTC's goals are to strengthen the scientific foundation of toxicological decision making and to facilitate the continuous improvement of the toxicological

toolbox.¹ Guiding principles of the EBTC include transparency, objectivity, and consistency (Hoffmann et al., in press).

These evidence-based principles and approaches are being drawn and adapted from the more established field of Evidence-based Medicine (EBM) (Hartung, 2009). EBM pioneered the use of systematic reviews and similar structured, transparent approaches to assessing evidence concerning health care interventions. Innovative but sporadic attempts have been made

* Disclaimer: The individual authors do not necessarily represent the views or policies of their institutions nor endorse all statements herein.

¹ <http://www.ebtox.com>



Tab. 1: The four core sessions of the Evidence-based Toxicology Collaboration (EBTC) workshop on Evidence-based Toxicology for the 21st Century: Opportunities and Challenges

Held January 24-25, 2012 in Research Triangle Park, North Carolina, USA*

Session Topics	1. Pathway Validation	2. Case Studies	3. EBTC Priorities	4. EBTC Governance
Presenters	Richard Judson (US EPA)	Ellen Silbergeld (Johns Hopkins University)	Suzanne Fitzpatrick (US FDA)	John Fowle (US EPA, retired)
		Daland Juberg (Dow Agrosiences)		
		Patricia Harlow (US FDA)		
Invited Discussants	Douglas Wolf (US EPA)	Robert Chapin (Pfizer)	Tim Pastoor (Syngenta)	Roberta Scherer (Johns Hopkins)
	Ed Carney (Dow Chemical)	Kim Boekelheide (Brown University)	Olga Naidenko (Environmental Working Group)	Rashid Shaikh (Health Effects Institute)
	Grace Patlewicz (DuPont)	Richard Judson (US EPA)	Michael Holsapple (Battelle)	Dennis Devlin (ExxonMobil)
				Andrew Rowan (Humane Society International)

* The full workshop agenda, which included introductory and concluding presentations, is available online at <http://www.ebttox.com>.

to apply these approaches in toxicology (e.g., Guzelian et al., 2005; Navas-Acien, 2006; Griesinger et al., 2009). The EBTC has taken on the challenge of establishing evidence-based toxicology (EBT) in a more organized and sustained effort.

To jumpstart its efforts, the EBTC held a workshop on “Evidence-based Toxicology for the 21st Century: Opportunities and Challenges” at the U.S. Environmental Protection Agency campus in Research Triangle Park, North Carolina, USA on January 24-25, 2012. The backbone of the workshop was a set of four sessions, each kicked off by one or more formal presentations, followed by invited discussants, and then general discussions (Tab. 1). The full workshop agenda, which included introductory and concluding presentations, is available online.²

The program reflected a number of the EBTC’s priorities. First, the EBTC is seeking to apply evidence-based methods to assessing the performance of emerging pathway-based methods consistent with the National Research Council (NRC) report on “Toxicity Testing in the 21st Century” (NRC, 2007). In keeping with this priority, workshop Session 1 featured a presentation by Richard Judson on the validation of high-throughput pathway-based assays (Judson et al., 2013) and Session 3 featured a presentation by Suzanne Fitzpatrick on a potential role for EBT in 21st century validation strategies (see Section 5.1). Similarly, two of the three presentations in Session 2 addressed fundamental questions relevant to establishing the scientific basis of pathway-based assays – one by Daland Juberg on the challenge of distinguishing adverse responses in these assays from those that are adaptive or compensatory (see Keller et al., 2012), and one by Patricia Harlow on the process of confirming that biomarkers used in these and related assays truly reflect the biology of interest (see Section 4.1).

A second EBTC priority is establishing an administrative structure and operational procedures that not only facilitate the work of the EBTC but also reflect core EBM/EBT principles, including transparency, continuous improvement, and volunteer inclusion. This priority was addressed in a presentation by John Fowle in workshop Session 4 (see Section 6.1).

And finally, Ellen Silbergeld, who gave a presentation in Session 2, provided helpful background information on evidence-based approaches and early attempts to apply them to EBT (Silbergeld and Scherer, 2013).

Given the novelty of evidence-based approaches in toxicology, the workshop program included ample time for both commentaries by invited discussants and discussion among workshop participants. To facilitate this commentary, the primary speakers were asked to circulate their presentations – in the form of white papers or Powerpoint slides – in advance to the invited discussants. The discussants were free to react to the presentations in their sessions or to discuss related issues.

The papers associated with the formal presentations in each of the four workshop sessions either appear elsewhere in this issue of *ALTEX* (Judson, 2013; Silbergeld and Scherer, 2013) or are summarized below, except the one associated with Daland Juberg’s presentation, which was already in press at the time of the EBTC workshop (Keller et al., 2012). A science writer was contracted to take the lead in summarizing the invited commentaries and open discussions for the proceedings, except where noted otherwise below. The invited discussants and the commentators were given an opportunity to review and offer edits to the draft summary prior to its completion.

² <http://www.ebttox.com>



2 Background information on Evidence-based Medicine and Evidence-based Toxicology

Evidence-based approaches have strengthened the scientific foundation of decision making in clinical medicine and health care by providing a structured framework for assessing the evidence bearing on healthcare questions. Moreover, such critical appraisals of past studies inevitably encourage improvements in the prospective design and reporting of new studies. Evidence-based tools are expected to have a similar impact on toxicology when appropriately translated from the medical context.

The primary tool of EBM is the systematic review, which includes a variety of steps: framing the question to be addressed and deciding on how relevant studies will be identified and retrieved, which studies will be excluded from the analysis, how the included studies will be appraised for their risk of bias/quality, and how the data will be synthesized across studies (e.g., meta-analysis). Such reviews also reflect EBM's hallmark tenets of transparency, objectivity, and consistency to the maximum extent possible. In addition, systematic reviews provide a convenient way for interested stakeholders to gain a condensed snapshot of the key literature and findings on a given subject.

Although EBM has earlier historical antecedents, its rise as a distinct discipline is usually credited to the work and advocacy of Scottish epidemiologist Archie Cochrane (e.g., Mayer, 2004). The term "evidence-based" was coined by Gordon Guyatt in 1990, and "evidence-based medicine" first appears in the medical literature in 1992 (Guyatt et al., 1992). The Cochrane Collaboration, named in honor of Archie Cochrane, was launched at Oxford University in 1993 to promote evidence-based reviews of the clinical medicine literature. By 2011, the Cochrane Collaboration had more than 28,000 active contributors in more than 100 countries, and the Cochrane Library³ contained more than 4,400 systematic reviews (see commentary by Scherer in Section 6.2).

The translation of evidence-based approaches from medicine to toxicology is already underway, at least at the conceptual level, but this process is only a decade old and still in the formative stage. Guzelian et al. (2005) coined the phrase "evidence-based toxicology" (EBT) and noted its promise in assessing the evidence that specific chemicals cause specific health effects in humans. Around the same time, Hoffmann and Hartung (2005) noted the potential value in translating evidence-based assessments of diagnostic measures in medicine to assessments of test methods in toxicology. Hartung and Hoffmann went on to further elaborate the conceptual underpinnings of EBT (Hoffmann and Hartung, 2006; Hartung, 2009) and coordinate the first international conference on EBT, held in Italy in 2007 (Griesinger et al., 2009).

Hartung later founded the EBTC – the collaboration devoted to advancing EBT – with several partners.⁴ For practical purposes, separate Steering Committees were established in the

US and Europe. The EBTC traces its beginnings to a 2010 workshop in the United States on "21st Century Validation for 21st Century Tools," which featured a session on the potential for evidence-based approaches to assess the performance of a new generation of non-animal test methods (Hartung, 2010). The EBTC formed in the ensuing months and held a kick-off meeting on EBT as a satellite to the 2011 Society of Toxicology annual conference in the US. The meeting familiarized the primarily US-based audience with the basic concepts and promise of EBT (Zurlo, 2011). It was well attended and generated enthusiasm for pressing forward to develop EBT approaches. The tasks to be tackled included identifying priorities, establishing work groups, developing a governance structure and work processes as appropriate, and engaging interested stakeholders in the process. It was decided to jumpstart these efforts by holding an open workshop.

The March 2011 kick-off event was replicated at the 2012 EuroTox conference for the European branch of the EBTC (Hoffmann, 2012).

Given its genesis at a workshop on validation of new methods, the EBTC has retained a keen interest in applying evidence-based approaches to assessing test method performance.

The launching of the EBTC is timely, as the toxicology literature increasingly invokes EB-related themes or practices such as transparency in decision-making (Schreider et al., 2010), systematic and transparent reviews of evidence (Sutton et al., 2011), synthesis of types of evidence to establish causal inference (Adami et al., 2011), and assessment of bias/credibility (Conrad and Becker, 2010). We also see practical examples of the application of evidence-based methodology or terminology (Woodruff et al., 2011; Abhyankar et al., 2011; Maull et al., 2012).

Evidence-based approaches provide a means of critically appraising evidence in a manner that is transparent, objective, and consistent. By contrast, standard toxicological practice still includes narrative (and thus subjective) reviews, non-transparent weight-of-evidence approaches, inconsistent decision-making procedures regarding the assessment of the hazards and risk of individual compounds, and reliance on aged toxicological methods of unclear performance. These practices compromise decision making and retard innovation in testing methods. Innovation is hampered further by prevailing frameworks for assessing toxicological test methods (i.e., validation), which are time-consuming, expensive, and intrinsically biased towards traditional methods.

To remedy these limitations, approaches have been proposed that build on EBM methodologies (such as systematic reviews) and practices (such as those established by the Cochrane Collaboration⁵). Unfortunately, attempts at implementing these approaches in toxicology remain uncommon and fragmentary.

The North Carolina workshop provided a forum for EBTC members and interested stakeholders to offer and discuss toxicological and organizational priorities for the collaboration.

³ <http://www.thecochranelibrary.com>

⁴ <http://www.ebtox.com>

⁵ <http://www.cochrane.org>



3 Workshop Session 1: The validation of high-throughput pathway-based assays

This session featured a presentation by Richard Judson (see Judson et al., 2013).

3.1 Invited discussant Douglas Wolf

The first discussant was Douglas Wolf, who at the time was the acting director of the Environmental Protection Agency's (EPA's) Exposure Assessment Coordination Policy division in the Office of Science Coordination and Policy, which, in turn, is in the Office of Chemical Safety and Pollution Prevention. This division is responsible for the Endocrine Disruptor Screening Program (EDSP). Wolf has since become the Assistant Laboratory Director of EPA's National Health and Environmental Effects Research Laboratory, which is part of the agency's Office of Research and Development.

Wolf raised several points. He noted that the 1996 Food Quality Protection Act has language stipulating that EPA must use validated methods to determine if a chemical has the ability to disrupt the endocrine system. In the plan published on the EDSP website, EPA is proposing a phased process. Initially, the agency is using high-throughput methods to help inform its process of prioritizing chemicals for more detailed testing. The results of the high-throughput assays, together with data from the agency's exposure evaluation and other available information, are helping the agency to determine which chemicals should be tested initially.

Over time, EPA will begin to replace current methods with high-throughput assays, Wolf predicted. This will happen after the agency evaluates the high-throughput methods, is able to show that they do indeed predict the potential for disrupting the endocrine system, and can either validate them or build confidence in their use. In the short term, Wolf said it will be easy to replace current *in vitro* methods with higher throughput assays. In the longer term, Wolf predicted that the high-throughput assays will replace whole animal methods.

Wolf suggested that the EBTC consider focusing its validation-oriented efforts on the high-throughput assays that were purposely designed for robotic systems, because we do not have experience validating these higher throughput systems. Efforts to determine how to validate them would constitute the best use of the EBTC's time and resources.

Wolf said that the EPA's goal is to provide a level of confidence in the agency's testing methods that will assure their validity to all stakeholders.

Finally, Wolf pointed out that it is important to have cell-based assays to evaluate a chemical's impact not only on human health but also on the environment, such as effects on wildlife. Most of the focus to date has been on systems based on human cells, cell lines, and mechanistic considerations, mainly because the pharmaceutical industry has taken the lead in developing high-throughput assay systems. "We at the EPA are equally concerned with environment and wildlife as we are with human health, so all of those are going to be necessary." The research community should be thinking about what kind of high-throughput assays

would be useful for predicting wildlife effects, as well as the technical concerns that may prove important, such as how to run assays for cold-blooded animals, he said.

3.2 Invited discussant Ed Carney

Ed Carney of the Dow Chemical Company was the second invited discussant. He noted that some whole-animal-based assays for developmental and reproductive toxicity have been able to identify unique effects that scientists otherwise would not have been able to predict. In his experience, however, the observed effects are mostly seen only at doses that are orders of magnitude higher than human exposure.

Carney said that his own declining satisfaction with the utility of whole-animal tests reinforces the need for a new paradigm. He agreed with Wolf that a prerequisite for moving forward with high-throughput pathway-based assays is a process whereby all stakeholders gain confidence in their utility.

Carney argued that the best place to begin using high-throughput cell-based assays is with prioritization. There is a strong need for increasing the capacity of doing such baseline assessments, given that there is only a limited capacity for conducting animal-based tests. The throughput (high) and the cost (low) of the pathway-based assays match the need, he said. Currently, prioritization is needed mainly for chemicals with very little data, so in effect, the bar is lower than for other applications such as risk assessment.

However, Carney cautioned that fitness for prioritization does not guarantee fitness for risk assessment. We need operating discipline to prevent inappropriate uses, he said. For example, the post-implantation rat whole-embryo culture (WEC) and the Embryonic Stem Cell Test (EST) are appropriate for embryo toxicity screening; he noted that the European Centre for the Validation of Alternative Methods (ECVAM) is careful to use qualifying language stressing that these cell-based assays are not valid as replacements for *in vivo* developmental toxicity tests.

A strong causal connection between altered biology in the high-throughput assays and adverse apical outcomes is needed to justify use in decision-making, he said. The strength of this connection currently varies with different types of cell-based assays. He offered the examples of the estrogen receptor and androgen receptor as ones in which both the pathways and the requirements for adverse outcomes are relatively well understood. The association is weaker for other assay endpoints. Case examples are needed to establish proof of principles, he said.

A practical challenge is to choose which assays to enter into validation exercises, he noted. Some are ready; some are not. Establishing an assay's suitability for a narrow intended purpose should be relatively straightforward. A harder challenge comes when moving from single assays to groups of assays that query pathways. Validation then should shift to evaluating the predictive value of groups of complementary assays, Carney asserted.

Carney stressed that it is also important to consider the nature of industrial chemicals. They often have very general functional properties that drive their use in commerce, such as surfactants in cleaning agents. This raises the possibility of confounding results, such as when a tested chemical denatures a target pro-



tein. Carney sees value in prioritizing chemicals into two broad categories – those exhibiting non-specific interactions, such as membrane alteration, and those known to have specific interactions, such as effects mediated by receptors.

3.3 Invited discussant Grace Patlewicz (with Richard Becker and Ted Simon)

The third respondent to the Judson presentation was Grace Patlewicz of the DuPont Haskell Global Centers for Health and Environmental Sciences. She opted to submit a written text for the proceedings in lieu of having her oral remarks summarized by the science writer. The text (comprising the balance of Section 3.3.) provides recommendations focused on the paper prepared by Judson et al. (2013) and is a summary of a longer document, “ACC Perspectives on Validation of High-Throughput Assays Supporting 21st Century Toxicity Evaluation,” prepared by the American Chemistry Council (ACC) Computational Profiling Workgroup, which she co-chairs with Richard Becker of the American Chemistry Council.

Everyone recognizes that the means by which data are generated and translated into information for the purposes of chemical regulation and risk assessment is undergoing a massive transformation. There are many drivers for this shift, including animal welfare considerations, advances in scientific techniques to rapidly screen substances for biological activities, and the large number of substances that exist in commerce for which toxicity information can vary to a considerable degree. Advances in high-throughput technologies, including *in vitro* cell-based assays and toxicogenomics, show considerable promise in changing the manner in which toxicity testing is performed in the future. However, the framework and context in which these types of data are evaluated and interpreted for regulatory decisions also will need to transform. The current validation process to develop scientific confidence in new methods to predict toxicity, as described by bodies such as ECVAM, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), and the Organisation for Economic Cooperation and Development (OECD), also will need to adapt to encompass the scenario of batteries of *in vitro* assays being used to evaluate a given regulatory endpoint, rather than the current framework of a single assay replacing a single *in vivo* test protocol.

To develop scientific confidence in high-throughput assays, the approach described in the Institute of Medicine’s (IOM’s) *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease* – namely analytic validation, qualification, and utilization – warrants consideration. Analytical validation entails analyses of available evidence on the analytical performance of an assay. Qualification requires the assessment of available evidence on associations between the measured biomarker response and adverse effects. Utilization requires contextual analysis based on the specific use proposed and the applicability of available evidence to this use. For high-throughput assays such as those in the ToxCast™ program, these steps could be adapted as: 1) Analytical validation, a consideration of the performance of an assay or suite of assays; 2) Qualification, an assessment of the association of the assay with a molecular initiating event, key event, or biomarker within the mode of

action (MOA); and 3) Utilization, a contextual and weight-of-evidence analysis of a specific use of quantitative assay results based on all available evidence.

With respect to analytical validation, one particularly difficult issue noted by Judson et al. is how to validate proprietary assays or those that require extensive robotics. A pragmatic approach for this validation could be to focus on the core basis of the assay and conduct either a non-proprietary assay that is complementary to the proprietary assay or to conduct one or more of the high-throughput assays in a low throughput mode. Neither proprietary nor throughput considerations should be used as impediments to robust scientific evaluation. Heuristic methods will also be important, as will consideration of the combined assay results – chemicals with established activities can serve as performance standards, and lack of consistency between assay results for these chemicals should be a red flag. Cross-laboratory testing may also be addressed by performance standards. This is a core element of validation that should not be avoided due to practical difficulties.

Qualification addresses how well the assay results reflect key events within the mode of action. Qualification will necessarily need to consider differences in potency and efficacy. It will be particularly challenging to establish the links between the assay results, the MOA, dose responses, and adverse effects. It is vital to understand how results of high-throughput assays relate to biological response pathways (Bhattacharya et al., 2011; Seed et al., 2005), especially how the quantitative result from an assay relates to the transition from an adaptive to an adverse response. With current understanding, it will be challenging to distinguish results that represent transient and homeostatic responses from results that reflect adaptive responses (that may be reversible) from results that are sufficient to cause an adverse outcome.

Utilization raises the question of what level of scientific confidence is needed for different purposes, whether they be screening, prioritization, hazard identification, or hazard prediction. Clearly, any evaluation should have the end application in mind (Judson et al.’s “use case”), rather than conducting a “validation” *in vacuo*. For example, applying high-throughput methods for priority setting would tolerate greater uncertainty than the same methods being applied to trigger a risk management regulatory decision. EBT, Bayesian approaches, or other methods may prove useful in integrating many sources of information to arrive at a quantitative weight-of-evidence-assessment. Furthermore, dosimetry and exposure are key in providing the appropriate context of how a given test concentration in a high-throughput test system relates to a real life exposure. Use of high-throughput assay results in lieu of traditional toxicity tests to support hazard identification or hazard prediction is perhaps the most difficult stage to address in terms of establishing what level of correspondence between assay results and key events is really needed.

Twenty-first century toxicology presents an exciting era for toxicologists, risk assessors, and researchers. Programs such as the U.S. EPA’s ToxCast™ exemplify how new technologies can be exploited to address the challenges of risk assessment. The challenge remaining will be validation/evaluation and gaining an understanding of the strengths and limitations of these types



of technologies for specific uses. Scientific consensus among regulators, regulated entities, and stakeholders will be needed in order to engender confidence in the use of high-throughput prediction models and results for decision-making. Consensus may be achieved by application of a scientifically sound validation framework coupled with transparency (data and algorithms) and responsible communication. Appropriate peer review, communication, and outreach are of paramount importance to the successful implementation, acceptance, and use of these new technologies by all stakeholders.

3.4 Open discussion

After the floor was opened for questions, John Fowle, a member of the US EBTC Steering Committee, asked if there was a way to inspire companies like Dow or DuPont to share their *in vitro* and *in vivo* discovery data on adverse outcomes to help improve Quantitative Structure Activity Relationship (QSAR) models.

Carney responded that his company (Dow) was exploring new approaches to bringing novel chemicals to market and integrating QSAR data. “We are quite interested in working with government agencies, particularly EPA, to look at different ways of getting new products approved,” he said. He said that Dow is currently involved in a project associated with greener and less hazardous substitutes, which uses some newer tools, such as *in vitro* assays and zebrafish embryos, as well as computerized structure activity relationship (SAR) analyses of structurally similar analogs for which existing data might already be available.

Carney said that this project might eventually become a prototype for a different approach to producing the Pre-Manufacture Notifications that companies must submit to EPA to comply with the Toxic Substances Control Act. In place of guideline data for, say, fish acute toxicity, it might be possible to substitute a zebrafish embryo screen and SAR analysis, he suggested.

Wolf sketched an interactive approach to testing that could inspire chemical and pharmaceutical companies to share their discovery data. He drew an analogy to companies developing open-source software applications. Many such applications are available that people can download for free; users can offer suggestions for improvements. The software’s originator then can capitalize on the suggestions to make a profitable product. He argued that this kind of interactive approach would help amass important data for improving QSAR models.

Errol Zeiger, an independent consultant, argued that data used to make decisions about the tests or test programs should be available to other people in case they want to do their own analyses. He added that high-throughput assays will be a success or failure based on the individual tests used as screening tools, where the false-negative rate plays a crucial role, but can rarely be assessed.

Judson pointed out that every false-positive screening test result can force a chemical company to spend half a million dollars to conduct a full battery of tests to prove that its chemical is actually not problematic.

Thomas Hartung of Johns Hopkins University commented that EBT is a new toolbox that needs to be evaluated. The beauty of the evidence-based approach is to transparently define a process before you do it, thereby increasing objectivity, he stressed.

Mel Andersen of the Hamner Institutes for Health Sciences said that the most important information is what are we doing and why are we doing it – in other words, in how a problem is defined. He asked Judson if all of the assays being used to identify modes of action were really required to effectively identify them.

Judson replied that for any given context, there is a “sweet spot” – not too many assays and not too few. He also noted the importance of assessing a test’s fitness for purpose in terms of balancing its sensitivity and specificity to meet the test objective. Screening tests should be relatively sensitive and specific, although he acknowledged that it is hard to have both; cost is also a factor. In any case, the purpose is not to say that a chemical is likely to be, say, positive in a guideline assay, he stressed.

Carl Westmoreland of Unilever, chair of workshop Session 1 and a member of the EBTC’s European Steering Committee, commented that transparency was an issue that cropped up more than once in the discussion. Ellen Silbergeld of Johns Hopkins University explained that transparency has acquired a particular meaning from its association with the Cochrane Collaboration for evidence-based medicine and health care (EBM/HC). That definition is incompatible with the idea that information can be shared on a “need-to-know” basis, she said. As defined by Cochrane, transparency is an absolute condition, Silbergeld stressed. Because the concept of transparency is fundamental to EBM, Silbergeld contended that it must also be a key tenet of EBT.

Wolf responded that an evidence-based assessment of test method performance requires access to sufficient information to make a determination about the relevance, reliability, and fitness for purpose of a given assay or set of assays that are being incorporated into a test method. This begs the question of how one determines that the amount of information is sufficient to make a scientifically defensible determination that a given test method is validated or validate-able; the answer may need to be determined on a case-by-case basis.

Roberta Scherrer of the Cochrane Center at Johns Hopkins University pointed out that her center’s focus on transparency is driven by the goal of ensuring that all systematic review findings can be replicated. The organization has a publicly available protocol for how to synthesize evidence to ensure transparency.

Patlewicz suggested that the approach the OECD is using for characterizing QSAR methods could be considered a useful framework for considering the level of information required to establish a given assay’s suitability for purpose. She lauded the framework for its discussion of measures of robustness and how it defined its domain of applicability.

Maurice Whelan of the European Commission noted that ECVAM came up with the concept of performance-based test guidelines to help ensure that the agency was not “embarking on an endless series of tests.” The guidelines come into play in instances where two tests have essentially similar components that are intended to determine the same biological outcome. Once performance methods have been defined to establish reproducibility, capacity, and accuracy, they can be applied to all of the methods aimed at that outcome. Performance standards also can be used to help guide test developers on how to imple-



ment different versions of essentially the same assay on different platforms, Whelan said.

Performance standards may prove very helpful in the early stages of implementing high-throughput testing methods, Wolf commented. They hold promise for targeting peer-review to whatever a given assay is intended to accomplish, as well as what a validation process is intended to achieve. Because the assays are being used as part of a constellation meant to help provide decision support, they will continually evolve and be replaced. Performance standards could avoid the need to re-validate every time a method or assay is replaced by an improved version. Such an approach also would allow for transparency, he said.

Session chair James Freeman of ExxonMobil characterized what we're talking about with EBT as a shift from validation to "fit-for-purpose." We're all steeped in the validation model, but maybe there is a way to get out of the validation box and move to fit-for-purpose, he said.

Carl Westmoreland noted the difficulty in judging when an assay has relevance for a toxicity endpoint or pathway of interest. The issue is tied up with the relevance of the test itself, the cell line you are looking at, and the concentrations that you are testing. He invited participants to weigh in on how relevance could be judged in such cases.

Judson responded that relevance is where test assays will pass or fail. Today we can run a set of assays with a set of chemicals for which we have *in vivo* data, and we can find correlations. The biology makes sense. The difficulty is to step beyond that to suggest that correlation is causation. It is hard to support a claim that the output from many of the new technologies with different cell types, such as pathway modeling, virtual tissue modeling, orthogonal assays, and sandwich assays, really is telling you something definitive about what is going to happen in the animal, he said.

Traditionally, toxicologists are trained to think of "relevance" mainly in terms of the ability to predict a type of gold standard, Hartung said. Such a gold standard does not always exist. However, a crucial part of validating a given test is assessing its scientific basis. It is a type of relevance that the current validation paradigm does not exploit, yet it is exactly the kind of relevance needed to establish novel assays. We need to model pathways that are shown scientifically to be relevant for a hazard we want to study, he said.

This is where the objective and transparent assessments possible with EBT can come into play, Hartung continued. When we can show that a scientifically sound pathway has been demonstrated, then we have the criteria we need to justify the use of an assay or group of assays. This is the beauty of a different type of system. We should think not about being empiric by reproducing data from another type of test system, but by being scientific by demonstrating that we reflect the science as we know it with the tools of science and by the scientific method, he said.

Whelan pointed out that the process everyone struggles with is chemical selection. Because there are so many competing criteria for selecting reference chemicals, he argued for the value of having many sets of reference chemicals to use to challenge

high-throughput tests in terms of their predictive capacity. He asked about how the tools of EBT might be used to facilitate the process of determining what types of references toxicologists should strive for – and how to select chemicals to use as references.

Some of the questions that people have posed may be ones that EBT is not designed to answer, Silbergeld said. She likened EBM to a court of law. EBM does not determine who is guilty and innocent. It provides the rules of process, such as which information is admitted into the discussion and to the judgment process. EBT, "if it is following that same path, will do the same. It provides a set of methods that allow you to reduce bias in terms of scanning the fact landscape and presenting the information in a fully transparent method so everybody can replicate the process – not the experiments – by which you identified the outcome by passing it through evaluatory filters. It is not going to say that this is more relevant than that. It will say that this is less biased information that can be utilized in reaching your judgment."

Hartung explained how EBT might aid in solving the problem Whelan posed about compiling a list of reference compounds. An evidence-based approach would be to define the process of compiling a list of reference compounds, to define the substances to be considered, the criteria to be applied. Before being executed, the process would be peer-reviewed. All the stakeholders should agree that this is a fair way to identify the substances. This is a way of describing a process that can be reconstructed, which makes sense to somebody on the outside, and which does not require the involvement of a "pope."

Becker pointed out that one of the areas that are going to be most challenging is prediction modeling. He said that he could readily see how one could construct a prediction model using the EBT approach. To create a model, you would take the elements of the assays, an element of exposure and concentration, and other relevant data and knowledge of the substances. Then evidence-based approaches could be used for assessing the quality and reliability of the data, and then they could be plugged into the model. "That, to me, is the process we should focus in on as a good case study to begin the discussion of the application of [EBT]," he said, concluding the session.

4 Workshop Session 2: Case studies

This session featured presentations by Ellen Silbergeld (see Silbergeld and Scherer, 2013), Daland Juberg (see Keller et al., 2012), and Patricia Harlow (see Section 4.1).

4.1 Biomarker Qualification at the US FDA Center for Drug Evaluation and Research⁶

Patricia Harlow of the U.S. Food and Drug Administration's (FDA's) Center for Drug Evaluation and Research (CDER) gave an overview of the process that her center uses to qualify biomarkers, which they define as objectively measured characteristics that are an indicator of biologic processes (normal or

⁶ This section is based on a paper prepared for the proceedings by Patricia Harlow.



pathogenic) or a pharmacologic response to a therapeutic intervention (Biomarkers Definitions Working Group, 2001). The goal of biomarker qualification is to improve the efficiency of drug development by achieving a consensus across CDER on the interpretation of biomarker measurements used in drug applications. Historically, biomarkers have come into common use in an unstructured manner as the result of many separate studies published in the scientific literature over many years. A formal process for the qualification of biomarkers should reduce the use of animals by eliminating redundant studies conducted by sponsors who seek biomarkers for common purposes, particularly for safety. Such a process is likely to be more efficient and transparent than qualifying biomarkers on a case-by-case basis.

The end result of CDER's biomarker qualification is a conclusion that the biomarker results within the stated "context of use" can be relied upon to have a specific interpretation and application in drug development and regulatory decision making. Although a biomarker (i.e., the substance or analyte being measured) can become qualified, this qualification is not equivalent to approval of a specific test or diagnostic device for performing the measurement. Different assays can be used to measure a single biomarker as long as each assay has been demonstrated to measure the same analyte, and each assay has been appropriately validated. Once a biomarker is qualified, it can be used in the qualified context in drug applications. However, the use of a qualified biomarker depends upon the absence of 1) serious study flaws in collecting data, 2) application of the biomarker outside the qualified context of use, and 3) any new scientific evidence that conflicts with prior conclusions.

CDER initiated the development of its qualification process with a pilot program that involved two submissions for urinary biomarkers of nephrotoxicity in rats. The first pilot submission made by the Predictive Safety Testing Consortium (PSTC) in 2007 resulted in qualification of seven urinary biomarkers in 2008. The second pilot submission made by the International Life Sciences Institute (ILSI) / Health and Environmental Sciences Institute (HESI) Nephrotoxicity Working Group in 2008 resulted in qualification of two urinary biomarkers in 2010. One biomarker was qualified in both groups. Based on experience with these pilot submissions, a formal biomarker qualification process was proposed and approved by the CDER in 2009.

The framework for qualification of biomarkers and other drug development tools (DDTs) is provided in a CDER draft guidance on DDTs released in 2010 (US FDA, 2010). This guidance discusses two types of DDTs, biomarkers and Patient Reported Outcome instruments, and it describes the process of working with CDER as well as a process for consistent scientific evaluation. An appendix to the guidance will be issued for each DDT qualification. This process involves publication of a notice of qualification in the Federal Register and posting of the qualification letter on the CDER website.

The guidance on DDTs does not discuss evidentiary standards for qualification. The evidentiary standards (the type and amount of data) needed to support qualification of a biomarker will vary depending upon the specific proposed context of use.

For instance, a safety biomarker whose proposed context of use includes multiple test animal species used for toxicology studies could not be qualified with data obtained only in rats. Furthermore, the evidentiary standards for biomarkers with a clinical context are expected to be different from those biomarkers with a nonclinical context of use.

Evaluation of biomarker qualification submissions involves CDER personnel who are common to all submissions, as well as personnel who are specific to that particular submission. The process of qualification within CDER consists of two phases: a consultation and advice phase and a review phase. The end result is that the biomarker is either qualified or not qualified in the stated context of use, and the review is considered complete. The qualification letter is posted on the CDER website.⁷ Further description of the qualification process is provided at the CDER website and in the paper by Woodcock et al. (2011). The biomarker qualification program is designed to support development of DDTs but, at the same time, to minimize burdens on CDER product review divisions, whose primary responsibility is the evaluation of data supporting applications for drug development and drug marketing.

Currently, eight urinary biomarkers for nephrotoxicity in rats have been qualified by the US FDA, as well as by the European Medicines Agency. The submissions for these nephrotoxicity biomarkers contained study reports for rat toxicology studies that were primarily dose-response studies for multiple nephrotoxicants, as well as some non-nephrotoxicants. Using pooled data, the diagnostic performance of these biomarkers was evaluated in comparison with currently used nephrotoxicity markers (blood urea nitrogen [BUN] and serum creatinine) using receiver operating characteristic curves. The reference standard was histopathology of the kidney. The data indicated that these urinary biomarkers could either outperform or add value to BUN and serum creatinine in detecting certain drug-induced kidney lesions (Dieterle et al., 2010a; Harpur et al., 2011; Hoffmann et al., 2010; Ozer et al., 2010; Vaidya et al., 2010; Yu et al., 2010). Some evidence supports the utility of these qualified biomarkers to detect injury in specific nephron segments in the rat. Changes in urinary kidney injury molecule-1 (KIM-1), clusterin, albumin, trefoil factor-3 (TFF-3) were associated with injury to the proximal tubule. Changes in urinary cystatin C, total protein, and β 2-microglobulin were associated with glomerular injury. Changes in urinary clusterin were associated with injury to the distal tubule. Changes in urinary renal papillary antigen-1 were associated with injury to the collecting duct.

These qualified renal biomarkers can be used to facilitate drug development, as illustrated by decision algorithms such as those presented in Dieterle et al. (2010b). If a drug induces renal histologic lesions in rats but has no effect on the levels of BUN and serum creatinine, the sponsor could evaluate whether appropriate novel urinary biomarkers are diagnostic of the renal injury induced by the drug in rats. If the biomarker signal correlates with the evolution and reversibility of the histologic lesion, the biomarker would be considered diagnostic of injury in rats and could be used in defining the no adverse effect level (NOAEL)

⁷ <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/default.htm>



and the starting dose of the drug for first-in-man clinical studies. If a human assay is available for the particular biomarker, the sponsor could propose to monitor the biomarker in the clinical trial. If the biomarker was not diagnostic of injury in rats, then the proposed clinical trial could be delayed or a higher safety margin (lower starting dose) would be needed to initiate the trial.

A search of the CDER electronic document database provides evidence that the qualified urinary biomarkers are being used in drug regulation. Documents were obtained with reference to the qualified urinary biomarkers in regulatory reviews and other documents finalized from 2006 through mid-July 2011. Based on the number of documents obtained for each biomarker, the documents referring to KIM-1, clusterin, or cystatin C were examined in more detail. The number of regulatory documents with references to these biomarkers increased after 2008, the year the biomarkers were qualified. Although references to these biomarkers were found in documents from all divisions within CDER, two divisions (Division of Cardiovascular and Renal Products and Division of Metabolic and Endocrine Products) have produced the largest numbers of documents with references to these biomarkers. The references to the qualified biomarkers were found not only in reviews but also in communications with sponsors, such as meeting minutes and advice letters. Therefore, these qualified nephrotoxicity biomarkers are being used in CDER regulatory activities.

In addition to the qualification of nephrotoxicity biomarkers, cardiac troponins were qualified for nonclinical use on February 23, 2012 based on a submission summarizing the publicly available literature and CDER's experience with the clinical and nonclinical use of cardiac troponins. Of the submissions still in the qualification process, one submission is nearing the end of the review phase, and fourteen submissions are in the consultation and advice phase. Some of these submissions are for clinical biomarkers. The experience obtained thus far is being used to refine the process of qualification by formalizing written policies and procedures within CDER. Hopefully, a refined qualification process, as well as the qualification and use of biomarkers, will facilitate efficient drug development of an increasing number of safe drugs.

4.2 Invited discussant Kim Boekelheide

Invited discussant Kim Boekelheide of Brown University, a member of the US EBTC Steering Committee, observed that the EBTC is intending to focus primarily on assessing alternative test methods, techniques, and tools for analyzing toxic effects, rather than considering the toxic effects of chemicals per se.

Boekelheide rhetorically asked attendees if they can identify an approach that the scientific community is likely to agree upon for taking the new kinds of information being generated via 21st century toxicity testing assays and translating those into safety prediction for humans. The key issues have been raised before in the *Toxicity Testing in the 21st Century report* (NRC, 2007) and other forums, Boekelheide said. For example, what do we do about metabolism and QSAR databases in this new setting? What are the platforms that are useful and informative? How do we factor in important information such as human genetic variability and epigenetics? What computational tools can be

used to generate systems biology pathway models based on the *in vitro* signals that we identify so that they can be understood in the context of the larger biology? How do we translate from the *in vitro* concentration that we are using to an expected *in vivo* exposure, and how do we make these predictions?

Boekelheide also asked attendees if they felt that EBT has a role in defining what these components and modules are, as well as what the process is. Does EBT provide a first step in framing the structure that helps us find a way to make these new assays work for developing information that we consider to be informative and relevant for humans? This is a different question than asking whether an assay works, he pointed out. It is pondering the structure required for all of the different kinds of techniques that we require in order to make extrapolation work. Is there a process we as a community can agree upon that culminates in a larger sense of what that structure needs to be, he asked.

The sources of information for setting up our analysis of these structures, as well as the particular assays that will support them, include scientific articles and databases, Boekelheide continued. In general, there are very few scientific articles published in this area, which raises the question of what to do in the absence of information from published sources.

One potential source, Boekelheide said, is the large databases that now serve as repositories for information collected by new testing methods; these are growing by leaps and bounds. If you publish an article that uses a microarray approach, most journals require you to dump all of that microarray data into a publicly accessible database, he noted. The question is what to do with the huge amounts of freely accessible data now available. Is it the job of an evidence-based toxicologist to use that kind of information, he asked. If so, who gets paid to make use of what and how is it used?

Another issue Boekelheide raised is the fact that, unlike the medical field, there is currently very little commonality for collectively assessing information in toxicology. Every published paper is likely to use a different cell line, a different platform, a different array approach, and/or a different species from which the cell was derived, he lamented.

One possibility for integration is to take the pathways approach, Boekelheide said. Biological pathways and toxicity pathways are common across species, although there will of course be differences in details. But this interpretation relies on an assumption that he predicted will be Sisyphean to prove, that pathways will be the same and coherent.

Another key issue is how to ensure that the evidence-based approach will provide the appropriate level of detail to be informative about the information being generated. He viewed the case studies presented in the workshop session in question as lacking detail about some important data, such as the quality of the messenger RNA, which is going to strongly influence the signal integrity that one gets out in these platforms, he said.

Boekelheide also asked participants to consider whether evidence-based approaches can serve to guide the field prospectively as it designs new assays and tools, rather than simply be applied to the assessment of previously generated data.

Boekelheide concluded on a hopeful note by suggesting the evidence-based approach may allow bioinformaticians and stat-



isticians to systematically mine the large amounts of microarray information now in the public domain. This may allow the information to be manipulated, integrated, and understood in ways not previously possible.

4.3 Invited discussant Robert Chapin

Robert Chapin of Pfizer – a member of the US EBTC Steering Committee – speculated about how EBT would look in practice. One area where it may have promise is in comparing assays such as those for estrogenicity reporters. The existence of only 5 or 10 estrogenicity reporter assays suggests that there is a reasonable “substrate” for an EBT kind of approach for, say, identifying performance criteria, he said.

Because Chapin works in developmental toxicology, he said he has hoped EBT will be able to have an impact in that field. However, he noted that in the last 10 years, only 10 to 15 papers have reported on purported improvements on some variant of assays using zebrafish, whole embryos, and/or stem cells. This may violate one of the fundamental operating criteria of evidence-based approaches: that there should be a sufficient number of papers out there that all look at the same method and come up with x, y, or z performance characteristics to allow them to be compared, he said.

In this regard, Chapin asked Scherer (of the Cochrane Collaboration) how evidence-based medicine handles situations where there are not enough papers using the same method. She noted that Cochrane has an “empty” category; that in itself is a finding because it says that there is no evidence for a given question. She also pointed out that the Cochrane process evaluates medical evidence for heterogeneity. If the amount of heterogeneity is deemed critical – which she thought could be analogous to the situation Chapin described – the process stipulates that the data cannot be combined analytically. However, the Cochrane process does allow for a “narrative” review in such situations, whereby the results from the studies can be presented thematically.

Many of the papers that scientists may want to assess together will take the form of gene analysis reports, Chapin continued. Maybe one of the most valuable contributions that an EBT group could provide would be to encourage statisticians, modelers, and bioinformaticians to dive into the data and analyze them for commonalities. They may be able to find a new common denominator across studies that initially appear to have none.

Chapin concluded that the weakness of the evidence-based approach for some areas of toxicology is that there may not be enough similarity in the reports in the literature to start the process, a concern mentioned by Boekelheide. To get around this potential stumbling block may necessitate having experts comb through the huge volumes of available array data to bring commonalities to light.

4.4 Invited discussant Richard Judson

Richard Judson of the U.S. EPA, a member of the US EBTC Steering Committee, was the third and final discussant. He acknowledged that coming into this meeting he did not perceive a difference between EBT and validation. He said that the meeting had made clear to him that validation could be one of many focal points of EBT.

Judson discussed his thoughts on how an evidence-based approach might be used to develop a prioritization program like the U.S. EPA’s EDSP, a program that Wolf had discussed earlier (see Section 3.1). It would start with the hypothesis that chemicals to which people, fish, or frogs are exposed trigger perturbations to specific pathways. Because we believe that these pathways could lead to adversity, we want to test these triggering chemicals sooner rather than later. The next step is to look for evidence linking the pathways to perturbations or adversity. An evidence-based approach can be used to evaluate this. However, you need to make certain that the exposure is relevant and that the pharmacokinetics are evaluated, he said. Only if the exposure, pharmacokinetics, and pathways come together in an appropriate way can you get all the way from exposure to adversity.

There is currently a paucity of data for evaluating this, Judson said. We do have exposure measurements, some incidence information about what is in the nation’s waterways and air, some chemical use information, and details about how much is manufactured. We have some information about fate and transport properties, and we can ask if they are similar to chemicals that we already know a lot about. We also know that some of these chemicals have endocrine effects. These are all classes of information that we can throw into this process, he said.

Many datasets are available for use as sources of unbiased data, such as the ACTOR and PubMed, Judson said. EPA’s Integrated Risk Information System (IRIS) assessments for priority pollutants dataset, which constitute the gold standard for evaluating chemical toxicity, may be amassed via an evidence-based process, he noted. However, because we have at least 10,000 chemicals that we need to deal with, the IRIS would not be usable to evaluate them quickly enough because the tests take so long to be completed.

Judson’s comments led to a discussion about EBT’s utility that many conference participants said they found helpful. Chapin observed that, with EBT, scientists could evaluate, for example, whether a method does a good job of identifying *in vitro* biological activity, and if the findings are transferable across labs. It could be used as a process for evaluating methods, he said.

4.5 Open discussion

Thomas Hartung explained that evidence-based approaches are a certain toolbox that can be applied, in principle, to any type of scientific question. It is a way of condensing information with criteria that are transparent, objective, and explicit. It requires tools that have not yet been developed in the context of the new toxicology. So EBT could address any question in toxicology – a method evaluation, condensing information on a given substance such as arsenic, a medical treatment problem in clinical toxicology, etc.

One thing that makes EBT different is that the question is very clearly defined, he explained. Usually it is a limited, very precisely defined issue for which all relevant studies are collected. EBT is also different in that the process, not the content, is in the foreground. It is about constructing a process that is judged to be the best one possible at the time it is created. Ideally, the process itself is peer reviewed in advance to ensure that it is sound.



EBM also involves tools that are not generally used in toxicology, Hartung continued. These tools are used to address the quality of the data. It also includes formal mechanisms for condensing information.

Finally, there is a mechanism for making the information available so it is considered to be a resource of highest possible quality. Hartung expressed his hope that, in time, the EBTC would achieve a quality similar to that of the Cochrane Library,⁸ which is known to use a rigorous process to ensure that all of its data is of very high quality. In this way, EBT holds promise for providing a higher quality than journal peer-review, which can vary depending upon who is tapped as a reviewer, Hartung said.

What EBT is can change over time, just as EBM has changed and expanded over the years to become Evidence-based Health Care (EBHC), Hartung said.

Silbergeld reiterated an analogy from her presentation, comparing EBT to a court of law. There are rules of evidence that guide how information can be presented to the court. Rules of evidence do not tell the judge what to decide, she stressed. What rules of evidence do, and their value, is in imposing order on a chaotic world of things that present themselves as facts.

The process begins with coming up with some criteria as to what are relevant, perhaps some key words, which would inform how you would extract the information in question from the databases where it can be found. Then you amass all the available data – say, 1260 studies on arsenic. From there, you apply criteria that determine whether or not you will go any further with some of the elements that came up in your search, perhaps study size, a specific type of cell, etc. The end result is your collection of evidence.

Next, you assess the amassed material for aspects that increase your confidence in the findings. The evidence, with all its bumps and shining stars, is presented in the most transparent way possible. You want to have a way to compile the output, which could take many different forms, such as an odds ratio or a magnitude of change over a range of exposures, etc. This gives you a way to weigh the evidence, she explained. Then and only then can experts come in and make a judgment about what the evidence suggests.

Scherer added that the Cochrane Collaboration does not provide guidance as to how the evidence amassed by the process is to be used. It is up to the healthcare providers, insurance companies, consumers, and other interested parties to look at the evidence in the context of a specific patient and decide whether the evidence can be appropriately applied. “We do not make the judgment; we simply say here is the evidence,” she explained. The evidence does include conclusions about its reliability and the confidence that the Cochrane Collaboration has in the results.

Silbergeld stressed that EBT would in no way replace the wisdom of everyone assembled at the workshop. “The expert judgment, experience, and wisdom of toxicologists and others involved in the endeavors we’re all sharing are infinitely valuable,” she said adamantly. What Cochrane does is separate the process of collecting the information from those making judg-

ments about it. In this way, Silbergeld argued that it is preferable to the current approach taken by the agencies charged with shepherding toxicology data.

Hartung added that a lot of this is about condensing information so that it can be digested by the scientific community. As an example, he pointed out that ToxCast data is particularly useful for identifying toxicity pathways. EBT is useful for comparing how the ToxCast data identifies such pathways with ones identified via other methods, such as the Hamner approach using case studies.

The initial utility of EBT will be to build confidence with the public and stakeholders that what you are doing is relevant for decisions that are going to have great impact, Silbergeld said.

Rashid Shaikh of the Health Effects Institute expressed concern that the term “evidence-based” could be perceived as off-putting, given that scientists already see their work as evidence-based. He said that he perceived utility for EBT but also noted that implementing it will have its attendant challenges. For example, he foresees the possibility of narrowing down the subject of interest so much that the result is a filter so fine that not much comes out of it.

Silbergeld responded that it is possible to explicitly relax a filter if it becomes clear that it is allowing too few things to pass through. She added that she became an adherent of EBT due to her recognition that adopting it could have a positive impact on the scientific process. Cochrane has improved the practice of clinical trials enormously, she noted. *The British Medical Journal* also published a document detailing its utility for environmental and observational epidemiology.

Hartung said that a validation study is for EBT what a multicenter, randomized clinical trial is for EBM. In other words, it provides the most valuable information, he said. EBT teaches its adherents about how to set criteria to identify and thus, indirectly, how to produce high quality data.

Chapin noted that what the group is groping toward is a process and trying to find what the process will look like for hauling together the pieces of data that will be allowable in discussing whether a test method is appropriately delivering what its developers say that it does.

Silbergeld stressed that framing the problem is key.

Errol Zeiger, a consultant, voiced his developing understanding that EBT can be applied to data regardless of whether it results from a new technology or an old technology. “We’re talking about how to deal with data to evaluate a situation, answer questions, and recommend decisions,” he said. Many workshop participants indicated their assent to these comments.

George Woodall of the EPA’s National Center for Environmental Assessment (NCEA) noted that his division is in charge of the IRIS assessments for priority pollutants, and many of the recommendations made during this session echo those made by the National Academy of Sciences for the IRIS assessment on formaldehyde. Those things are being integrated into IRIS, he noted.

Nicole Kleinstreuer from the EPA’s National Center for Computational Toxicology (NCCT) said that her center initially

⁸ www.thecochranelibrary.com



identified the assays used to produce ToxCast data by using an EBT type of approach. Those assays come from a huge range of different platforms, including complex cell cultures, human primary cells, zebrafish, and stem cells. Each platform is subject to rigorous quality control and quality assurance. In a way, we are really trying to build up our body of evidence, she said. Kleinstreuer works closely with the predictive modeling project, and she is excited to apply this kind of framework to it. We've already started to build our basis of evidence from both the *in vitro* and *in vivo* sides, she said.

Hartung noted that ToxCast is not a great case study for applying EBT because it has already been subjected to so much quality control. The important thing for moving forward is the commitment to be systematic, objective, and transparent, he said. We desperately need to be objective and discuss biases, he said.

Many of the tools developed for EBM will fit quite well with what toxicologists do, Hartung continued. Systematic reviews can be applied; data appraisals are now under development; meta-analyses haven't yet been done, but they're coming; and test assessment methodologies can help improve toxicology, he said.

Scherer pointed out that the area of research synthesis is quite a new field. Systematic reviews weren't around before the 1980s, she said. What you are struggling with is a method for research synthesis in your field. No one has done this before – that is why it is so difficult.

A part of Cochrane that many people do not know about is the existence of methods groups, Scherer said. They do studies on how to carry out the elements of a systematic review. For example, an information retrieval group goes through and develops search strategies. They figure out the best way to search for a particular kind of study. There is also a bias methods group looking for potential biases. Another group focuses on diagnostic test accuracy, and one focuses on statistics. "This is an exciting time for toxicology because you're right on the edge of developing these methods," Scherer said.

Any kind of question can be addressed using an evidence-based approach, she reiterated. The question is critical. It is the foundation for how you conduct your search and how you appraise the studies that are going to be included in your review.

Mike Holsapple of Battelle, who is also a member of the US EBTC Steering Committee, pointed out that the draft EBT mission statement suggests that people will use the EBT method both to sort through the available data and to judge it to facilitate robust decision making, which is at odds with the Cochrane approach. James Freeman, who chaired this workshop session and is a member of the US EBTC Steering Committee, pointed out that, historically, toxicologists did everything from inventing the tests to running them and creating risk assessments. Toxicologists are used to evaluating their own work, he said. Silbergeld noted that some kind of division of labor is likely to be useful, and it seems to have served the Cochrane Collaboration. Hartung said that once the evidence (or the absence of evidence) has been mapped it can be used for a variety of pur-

poses, and different entities may interpret it in different ways depending upon their needs.

Martin Stephens of Johns Hopkins University, who is a member of the US EBTC Steering Committee, ended the session by showing a slide that was originally prepared for EBM by Scherer, and which he is translating to EBT. It gives an overview of the processes of generating evidence, as well as gathering and assessing published studies. Different people are involved with each step; industry, government, and academic labs generate the evidence; the data generators themselves can gather or assess it, as can members of the EBTC or others. Others can take it and integrate it into policy (if appropriate), including whomever in industry or government makes decisions. In the end, you have the application to public health.

Once an EBT assessment has been completed, it can also feedback and inform how new data gets generated and the types of studies that are conducted, Stephens said. In that way EBT can help not only to condense existing data but help shape future studies, he concluded.

5 Workshop Session 3: Potential Priorities for the Evidence-based Toxicology Collaboration

This session featured a presentation by Suzanne Fitzpatrick (see Section 5.1).

5.1 Twenty-first Century Validation Strategies – Can Evidence-Based Toxicology Play a Role? ⁹

The National Research Council report, "Toxicity Testing In the 21st Century: A Vision And A Strategy" (NRC, 2007), described a new vision and strategy for toxicity testing that would be based on human biology rather than animal biology and would be less expensive and time consuming. This vision also involves a strong commitment to the 3Rs – replacement, reduction, and refinement of animal use in experiments. Being responsive to this progressive scientific perspective would necessitate moving forward to develop, validate, and incorporate alternative toxicological test methods into the federal regulatory framework.

As stated in the NRC report, "[c]hange often involves a pivotal event that builds on previous history and opens the door to a new era." The publication of the NRC report itself was the "tipping point" for a change in toxicology, but validation of these new methods for regulatory use will be the critical component in ensuring the vision's success. Toxicologists have the unique opportunity to meet these challenges by looking at new approaches, new collaborations, and new ways to take advantage of 21st century technologies.

Current formal approaches to validation involve lengthy and expensive processes that require validating *in vitro* data from a single assay against *in vivo* data. These approaches are not relevant or even feasible for the new pathways and endpoints being measured with high-throughput and high content methods.

⁹ This section is based on a paper prepared for the proceedings by Suzanne Fitzpatrick and Abigail Jacobs.



Consequently, applying a “one size fits all” approach to validation is not conducive to the rapid incorporation of emerging science or technology into the regulatory decision-making framework. As new safety testing evolves, new approaches to demonstrating that a test is reliable and relevant for a particular purpose must also evolve.

An example of a new validation strategy is the FDA Drug Development Tool (DDT) Qualification Process (see Section 4.1). Qualification is a regulatory conclusion that, within the stated context of use, the results of an assessment with a DDT can be relied upon to have a specific interpretation and application in product development and regulatory decision making. Once a DDT is qualified for a specific context of use, industry can use the tool for the qualified purpose during product development, and FDA reviewers can be confident in applying the DDT without the underlying supporting data. The FDA DDT Qualification Program involves a “fit-for-purpose” qualification. It is an objective, science-based approach for evaluating the relevance, quality, and reliability of a biomarker based upon its intended use, for example, a test method used as part of a screening program or a definitive surrogate endpoint in a pivotal clinical trial. Details about the FDA’s qualification program are available online.¹⁰

There is a pressing need to develop and use a structured evaluative process for new toxicology tools similar to the DDT biomarker qualification process. This process would consist of uniform, objective, science-based criteria for systematically determining data relevance, quality, and reliability. Once novel, cutting-edge methods have been evaluated and incorporated in the toxicological toolbox, data on a chemical’s effects from new and existing methods from all relevant studies should be comprehensively reviewed, given appropriate weight, and integrated in a transparent manner. The resulting document would describe a robust, biologically plausible understanding of the mode of action of a chemical and the potential hazards and risks that exposure to the substance could pose to humans and to wildlife.

Regulators must ensure that their toxicological toolbox keeps pace with advances in science and technology. But regulators also must determine how much evidence is sufficient to judge that a new tool is qualified to inform safety decisions that potentially affect millions of consumers. There is a delicate balance between ensuring safety and avoiding restrictions on valuable products; this is a continual, unique, and demanding challenge for regulatory agencies. Any advances in validation must recognize these competing demands that regulators face.

It is now clear that *in vivo* animal studies cannot be the gold standard that we qualify new toxicology methods against. Regulators must determine the relevance of *in vitro* results to what occurs in humans rather than the concordance of that data to what occurs in rodents and other test animals.

A key aim of the EBTC is to advance 21st century toxicology by translating the principles and approaches of Evidence-based Medicine (EBM) to the evaluation of emerging toxicological testing methods. Structured reviews of existing evidence are a key feature of EBM, and translating this to toxicology could

give regulators the needed evaluative procedures to judge the quality of new toxicological tools through the use of EBT’s systematic reviews and meta-analyses.

Systematic reviews carried out in a regulatory context must have the ability to look at proprietary data in a transparent manner, while keeping the data confidential and addressing the open literature as well. Often regulators are criticized for not using data from the published literature. Applying systematic and meta-analysis approaches to the published studies in journals could assist regulators in incorporating these data into regulatory decisions.

There have been some objections to using EBT on the grounds that toxicologists already use evidence in assessing causation and reaching regulatory decisions. However, toxicologists should not let semantics turn them off to this approach. Systematic reviews would offer a complete and rule-based analysis of data with conclusions that are transparent enough to be reproducible.

What can the EBTC do to begin promoting the use of evidence-based reviews of *in vitro* toxicology test methods for making safety decisions for human exposure? A prerequisite is to develop a strong coalition of scientists from government, academics, and industry who are committed to working together to facilitate change. Agreement on a shared vision and a governance plan will give the EBTC a strong foundation to build consensus.

Single tests alone, even if highly sensitive or specific, can no longer provide an appropriate assessment of a chemical’s toxicological properties. Toxicologists will need to incorporate information from many diverse sources – omics, animals, tissue culture, engineered organ systems, *in silico*, etc., into integrated testing strategies. The EBTC should help to identify the critical questions that need to be addressed when assessing the safety of products. These questions undoubtedly will differ between regulatory agencies and even within the same regulatory agency. The EBTC could then develop draft evidence-based criteria for integrating data from different sources to assess the relevance to humans. How much data are needed to give regulators confidence in a new method? The answer will differ depending on where the method is incorporated into the regulatory paradigm.

The EBTC should identify both short and long term goals for addressing the issue of validation of new pathway-based methods. A good starting point might be the drafting of criteria for the validation of screening methods, where the levels of false negatives and false positives are not as critical to safety assessments.

Evidence-based methods can provide more than a kind of quality assurance for new methods. They also can be used to synthesize the toxicological data for a risk assessment. Test cases could be developed, comparing the results of evidence-based assessments with traditional risk assessments to see which gives the most clarity to assessing a product’s safety.

Workshops to encourage dialogue and consensus need to be ongoing to gain confidence in the new validation strategies. Regulators need to clearly articulate the level of performance

¹⁰ <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/default.htm>



needed to be adequate for the method's intended use. Draft guidance with notice and comment for public input can assure transparency in the new procedures. International collaborations with complementary research programs could help vet these new strategies worldwide.

Investments in advancing new validation strategies for 21st century methods can enable regulatory agencies to better protect and promote the health of people in the United States and throughout the world. Moving towards evidence-based approaches to validation is challenging but essential to catalyzing change that would allow us to take advantage of revolutions in science.

5.2 Invited discussant Tim Pastoor

Tim Pastoor of Syngenta anchored his remarks to RISK21, a multi-stakeholder project that he is co-chairing. Organized through the International Life Sciences Institute's Health and Environmental Sciences Institute (ILSI/HESI), RISK21 is focused on applying the new toxicology and exposure assessment tools for the 21st century, as defined by the landmark 2007 NRC report (NRC, 2007), to what he called a "risk context."

Pastoor expressed his belief that toxicology is at the brink of a huge change in how practitioners conduct risk assessments. The new technologies that are available, together with some of the new thinking and the impetus provided by the NRC report, are a mandate to change. "We have dug a very deep hole as toxicologists because everything is about hazard-based assessment," Pastoor said. He challenged the audience to begin thinking instead about an approach based on zones of safe exposure.

In place of the conventional approach of initially determining what effects a substance causes, RISK21 posits that the first consideration should be how much someone might be exposed to, Pastoor said. If human exposure is minimal to none and there is evidence that the toxicity is very low, one might do toxicity testing very differently.

Are we ready to move toward an approach that is exposure-based and focused on safety, Pastoor asked rhetorically. He emphasized the importance of using prior knowledge. The RISK21 group is exploring the idea of systematically assessing the body of prior knowledge using Bayesian network analysis, such as the probability within a class of toxicity potency.

Pastoor argued that the point estimates and calculations used to determine acceptable exposure, such as acceptable daily intake (ADI) and reference doses (RfD), suggest that there is a bright line between safety and lack of safety. The reality is less clear-cut.

RISK21 also calls for employing targeted *in vitro* testing, in conjunction with pharmacokinetic (PK) data, to narrow the range of expected toxicity, Pastoor said. In other words, you can use prior knowledge to make an estimate of what your toxicity values are likely to be, as well as what the exposure is likely to be. "Why don't you just do the studies that are necessary to refine that knowledge?" he asked workshop attendees.

RISK21 counsels using targeted *in vivo* studies, Pastoor said. He presented data from the RepDose database showing the distribution of no observable effect levels (NOELs) for 404 chemicals in mmol per kilogram of body weight per day. The toxicity

for the group varied over eight orders of magnitude, which he said "is what you'd expect – it happens in nature." This gives you a distribution to work with without doing a single toxicity test, he said. With some targeted testing, you have a good chance of understanding exactly where your chemical might be, he contended.

Pastoor noted the importance of comparing the exposure that people have to a chemical to the doses associated with toxicity. For example, if you know that only a relatively high quantity of a given chemical is likely to be toxic, and the quantity to which someone is likely to be exposed is much lower, you can feel pretty certain that it is unlikely to cause harm. In a situation where there is overlap between the dose likely to be toxic and the quantity to which people are likely to be exposed, you can refine your exposure assessment by targeted testing. You can also do the same thing for the hazard, he said. By doing so, you get a better idea of the gap between exposure and toxicity, he said.

"Let's get out of the rut," Pastoor concluded. The issue at stake is the protection of human health. "When we're talking about priorities, we're really thinking about how we use prior knowledge to understand things so we don't have to kill additional animals in a pointless effort to try to refine information that we really don't need."

5.3 Invited discussant Olga Naidenko

The comments of Olga Naidenko from the Environmental Working Group reflected a societal view of chemicals management and environmental stewardship.

"Despite the best work that toxicologists have been doing up until now.... we are still stuck with what many people feel are ... annoying and perhaps even intractable problems," she lamented. The process for screening endocrine disruptors is not very effective, occupational chemical exposure in the U.S. is not adequately assessed, and the best science is not being used, she argued.

Naidenko said that she agreed with a key point that Fitzpatrick made in her white paper: "The ability of researchers to develop and deploy biological profiling and high-throughput/high content methods has outpaced the ability of regulatory agencies to apply traditional method validation approaches for demonstrating relevance." In fact, Naidenko said she felt that "the ability of researchers to develop and deploy new chemicals has far outpaced the current ability of society to assure the safety of these chemicals for the long-term sustainability of the planetary society...."

Naidenko also agreed with Fitzpatrick about the need to build a broad-based coalition of scientists who are willing to work together to facilitate change and encourage the use of EBT, as well as on the need to develop uniform, objective, science-based criteria for systematically determining the relevance, quality and reliability of new test methods.

Naidenko made some suggestions about obstacles she perceived to EBT's implementation and how to overcome them. Anyone who has ever participated in a political coalition knows that actions of coalitions are strongly dependent on positions of their participants, which are in turn shaped by their experiences and values, she said.



The solution to these pitfalls is transparency about values and goals, Naidenko said. Open discourse about values ensures that regulators and stakeholders cannot hide behind “science made me do it” statements, she said.

Naidenko raised concerns about the white paper’s statement that, as new safety testing evolves scientists will develop “uniform, objective science-based criteria.” She pointed out that values and the perception of what is “safe” and what constitutes an “unacceptable risk” vary within society, between societies, and between different economic groups, such as workers, consumers, parents, the affluent, and those who are economically challenged.

Naidenko asked participants to consider “the balance of our goals in the economics/ecology continuum.” How do we handle risk-risk tradeoffs, she asked. For example, certain degrees of public health protection may entail the loss of jobs. This begs the question of what we can afford.

“We of course believe that our science is objective, but science is very much embedded in the society that produces the knowledge,” Naidenko said. Understanding the nature of knowledge is essential for helping us avoid errors in thinking, she said. In addition to providing useful insights into the issues facing toxicologists interested in using 21st century tools, the study of the sociology of science and progress may help us save time and money, Naidenko suggested.

Naidenko commented that what toxicologists are trying to do via EBT is inherently more difficult than what medical science has achieved through the Cochrane Collaboration. While Cochrane is focused on a better way to do good by curing patients, the toxicology collaboration may be seeking to determine the scientifically, ecologically, and economically acceptable level of harm, she observed.

We cannot do EBT successfully unless society jointly, transparently decides on its values along the economic-ecological continuum, she contended. She said she agrees with Fitzpatrick and Pastoor that we do want to use prior knowledge. We do want to study the literature on previous stumbling blocks, such as BPA, to determine what went wrong and what we can learn from those experiences. We don’t want to spend millions of dollars to end up with something that no one is satisfied with, she concluded.

5.4 Invited discussant Michael Holsapple

The third invited discussant, Michael Holsapple of Battelle and a member of the US EBTC Steering Committee, agreed with the points that Fitzpatrick raised in her white paper and said he would be presenting some of the issues that the EBTC will need to tackle in the future.

He emphasized that toxicologists have many topics of interest, but few have the potential for having a greater impact on the science of toxicology than the challenge of using 21st century tools, or Tox21c. He said he feels that if toxicologists do not embrace Tox21c and try to move it forward, the discipline is in danger of become increasingly marginalized.

The current “menu-driven” approach to toxicology provides too much information about items of low interest and insuffi-

cient data on high priority effects, Holsapple said. To amplify this point, Holsapple quoted from a recent editorial in *Science* by FDA Commissioner Margaret Hamburg (Hamburg, 2011), who was essentially suggesting that toxicology is somewhat stuck in the past, and we need to be moving forward. The FDA’s strategic plan calls for improving predictive models and modernizing toxicology to improve product safety, he continued.

Holsapple said that he was struck by how many people on the US EBTC Steering Committee observed that their understanding of EBT was enhanced by the talks at the workshop. There is still a real lack of consensus by the Steering Committee on what EBT is and should be accomplishing, he observed. “We’re really struggling with what transparency means, in terms of how we’re going to move this forward,” Holsapple continued. Like Fitzpatrick, he felt that listening to Silbergeld and Scherer describe evidence-based science had greatly enhanced his understanding of the discipline.

Holsapple said that he felt that members of the Steering Committee would benefit a great deal from an “EBT 101” type of introduction to the issue. This would help ensure that the ambassadors trying to advance EBT are all on the same page and articulating the same message, he said.

He said that he is also still struggling with the mission statement. He said his current understanding of evidence-based approaches is that the process involves amassing all of the evidence on a narrowly defined subject and going through it very systematically, using *a priori* defined criteria. The analysis involves making sure the evidence is packaged in such a way as to clearly show that it presents the best “look and feel” for what the evidence can present right now. “The important thing is that those involved in the evidence-based approach don’t make the value judgment; they don’t say ‘yes’ or ‘no’ or it’s ‘blue’ or ‘green.’ The evidence is handed off to someone else,” he summarized.

Holsapple believes the EBTC needs to develop a communication strategy to focus on ensuring that this evidence-based approach is understood. We have to become the ambassadors to champion that, he said. A key target audience is the public. “If we don’t bring the public along, they’re going to be blindsided, and they’re not going to be happy,” Holsapple said. He feels that the EBTC needs to engage in the same kind of outreach to other stakeholders because they are going to be very important to realizing the coalition’s goals.

5.5 Open discussion

Ellen Silbergeld complimented the panel of speakers for moving the group to think about the directions, actions, and opportunities for productive collaborations. She agreed with Holsapple that bringing along “the broader community and persons of interest will be absolutely critical.”

A wide-ranging discussion of issues related to exposure, dose, and safety ensued, mostly in response to Tim Pastoor’s invited commentary. Session chair Rodger Curren, a US EBTC Steering Committee member, asked participants to try to focus on the issues that the EBTC should be looking at in terms of validation and understanding how EBT methodology can be supplied.

Richard Becker pointed out that workshop participants seem



to have different visions for EBT. If we think very broadly, there are probably different applications of evidence-based approaches that we should be thinking about. We probably also should be thinking about moving forward within some kind of consortium to address these, he said. One component of EBT that is top of mind is 21st century tools and validation strategies for these tools, per Fitzpatrick's presentation. That may be one area the EBTC should consider focusing on, he said.

Becker said that he thought the group needed to look carefully at the decision paradigm for which those tools are being developed and for which they are intended to be applied. That has to be integral to the way that we evaluate the data to establish the scientific confidence in that tool, he said. He noted that Fitzpatrick said the FDA looks for the purpose for each methodology within a regulatory decision-making framework. Similarly, Becker observed that Pastoor talked about an integrated Bayesian assessment approach that involves looking at knowledge and targeted evaluation using different methods to provide the information needed to make the decisions.

Becker posited that the new technologies cannot be evaluated by divorcing them from the application for which they would be used. He therefore suggested that it is important to bring forward the tools, together with the data and the context and the decision framework within which they would be utilized.

Holsapple said that he felt of two minds about his response to Becker's comments. "I agree that we need to understand the context," he said. "I think that's essential." He said that he did not think it made sense to try to evaluate fitness for purpose for each of the many assays now available. Instead, he thought it made sense to get a fit for purpose for a series of assays.

On the other hand, he pointed out that, as he understood the evidence-based approach, it is a tool that can be applied to any question. You define the question up front and define the criteria a priori to answer any of these questions. This potentially broad scope notwithstanding, Holsapple argued for the EBTC to go on record as being involved in validation as one of its major activities. The other alternative is the more general goal of promoting the beauty of trying to conduct meta-analyses and very systematic reviews and bring evidence-based approaches into toxicology, he said.

These are two different paths, although they are not necessarily mutually exclusive, Holsapple said.

Fitzpatrick asked if EBT could be used to frame the question to see if evidence-based approaches could be used to evaluate fit for purpose. As an example, she said: "is this method, in this regulatory context, able to give us assurance that the answer is what we want?" If the question was framed tightly, it might result in a lot of EBT evaluations at first, but it would enable people to evaluate all of the evidence.

Holsapple said that he felt it would work as long as you were setting *a priori* criteria to address the question. However, Holsapple said that the answer of whether the fit was good or not would go beyond what the evidence-based process is designed to achieve.

Fitzpatrick said her idea was to use evidence-based approaches to analyze the question, and the regulator would look at the

synopsis to make a decision. "It would be a way to systematically review evidence by the same transparent process," she said.

Holsapple responded that one thing that was alluded to on the previous day was for the EBTC to conduct some sort of case study. In other words, to use the process to analyze some data that presents a daunting challenge and then have someone look at the data and make some sort of a determination as to whether or not the evidence-based process proved useful.

Silbergeld said that she did not think it made sense to try to tackle a daunting challenge. Pick something that a crisp question can be formulated to address, she urged. Ideally, this subject should be one that has been the focus of as many papers as possible so there is a lot of evidence to bring to bear. The topic also should be one for which researchers hope that it is possible to get somewhere by sifting through the evidence systematically.

Silbergeld again raised the analogy of a courtroom. "It's not like a grab-bag of different charges and different laws. It's a specific case for which evidence has been deemed to be admissible or not," she said. "If the charge were different, then maybe different evidence would be admissible." Formulating the question crisply is the key to success, she said.

Holsapple responded that, to him, picking a subject that was data-rich was part of what constituted a daunting challenge. "What's necessary for EBT to get some kind of momentum is to demonstrate that we as a community can apply it and get some meaningful information from it."

Fitzpatrick said that in drug development and qualification there is a "pre-EBT phase." We come in for a consultation before all of the evidence has been generated to determine what kind of evidence is needed to give the regulators confidence, she said.

Thomas Hartung pointed out that what the group collectively learned yesterday was that evidence-based approaches are a way of handling information and condensing it. The process is so good that scientists can rely on the end product without going through all of the supporting information in detail. The process ensures that the evidence has been evaluated in a credible process. By this we can help a lot of applied toxicology, he said.

Today we have heard some compelling calls for using the evidence-based process for evaluating candidate drugs and pesticides, Hartung added. It is clear that evidence-based approaches have the ability to condense information and ensure quality on all levels. What the presentations and discussions to date have shown is that one of the priorities for the EBTC is to sharpen the group's understanding of what the tools do and do not deliver, he said.

Hartung said that he was encouraged to hear that Fitzpatrick had already had positive experiences with applying some EBT-like processes to BPA. "It's most important that we now build from these types of cases," he said. He said that EBTC has some mechanisms to fund some meta-analyses and studies.

Hartung agreed with Holsapple's statement about the need to produce a document that clearly explains what EBT does so that people are discouraged from using the label for whatever they think is cool in toxicology at the moment. Holsapple said that, in addition to defining what EBT is, any document produced



should also make clear what it is not. Perhaps a flow chart could be used to show this, he said.

Hartung said that Naidenko's comments also made clear that, in addition to aiding the toxicological community, EBT may help with outreach to other stakeholder groups and the public. High quality, condensed information could appeal to many others and could be used to inform many processes, he said.

Sebastian Hoffmann of seh consulting, and a member of the European HTPC Steering Committee, said that from the European perspective, it should be a priority for EBT to be used not only to assess individual tests but also to develop testing strategies. He acknowledged that EU and U.S. priorities may differ.

Naidenko said that it may be important for the group to consider how to get stakeholder buy-in as early in the process as possible. She also observed that doing case studies and publishing them early on would be a foundation for moving forward.

Gillian Griffin of the Canadian Council on Animal Care pointed out that the 8th World Congress on Alternatives and Animal Use in the Life Sciences, held in Montreal last summer, resulted in a declaration on synthesis of evidence. She noted that conferees balked at calling it a declaration on systematic review, in part because people in the audience were concerned that we really did not have enough in the literature that would enable us to be able to do a systematic review. Because this is an iterative process, as we go forward, the information published in the literature becomes richer and we are able to apply it to do better systematic reviews. She said she was very excited that EBT also is moving in that direction and noted that a number of other fields are also doing so. Finally, she said she also felt that public engagement was key.

Doug Keller of Sanofi said that, as the EBTC defines itself and starts to tackle case studies, it's essential to establish itself as an unbiased organization. You don't want to be perceived as a group with an agenda to justify pathway-based methods or to replace animal use, even if that's the eventual goal that many people have. The Steering Committee's composition is helpful in that regard, but Keller stressed that the group's credibility will be established by how it proceeds.

Hartung said that an important aspect of evidence-based approaches is that they provide many tools to help practitioners avoid bias and to ensure that it does not affect the quality of the end product. This is, in part, because the process does not allow for manipulation, he said. Where manipulation cannot be excluded, the process makes its presence as evident as possible.

Silbergeld noted there are two kinds of bias. The first she defined as the National Academies kind of sociological bias, which is based on where you stand from and where you sit. There is also bias that is related to how studies are conducted and how data are collected. That is to some extent independent of the first type of bias, she said.

The whole reason for initially establishing *a priori* consensus-based definitions is to avoid bias, Silbergeld continued. Defining the characteristics or aspects of a paper that increase our confidence when present and decrease our confidence when

absent allows reviewers to have a common set of transparent reference points, she said.

Rob Wasserman of EPA's IRIS program said that his program is in the process of developing criteria for what constitutes acceptable evidence. He said that, to him, it seems that the first thing for the EBTC to do would be a comparison involving criteria that already exist, rather than a case study.

6 Workshop Session 4: Governance and work processes for the Evidence-based Toxicology Collaboration

This session featured a presentation by John Fowle. The following section (6.1) is a summary of that presentation.

6.1 Governance and Work Processes of the Evidence-Based Toxicology Collaboration

John Fowle, who recently retired from the EPA, discussed key points to consider as options for establishing governance and work processes for the EBTC in a white paper. The options are modeled after the governance and work processes of the Cochrane Collaboration and its approach to EBM.¹¹

The goals of the EBTC include fostering the development of a process for quality assurance of new and traditional tests for the assessment of safety in humans and the environment, as well as providing guidance on evaluating evidence from new and existing tests when assessing chemical safety. The key methods used by the Cochrane Collaboration include systematic reviews of relevant literature, such as inclusion/exclusion criteria for published studies and meta-analysis of data. These approaches can be applied to the EBTC, and in addition, the Cochrane Collaboration approach to governance offers a model for shaping the nascent organization's operations and policies.

The EBTC, in order to be successful, will need to reach the hearts as well as the minds of many individuals in many different areas of expertise to secure stakeholders' time and effort to systematically identify and review the toxicology literature. The organization also will need people to serve on teams to write reports of the findings; to work to capture the reports in databases and electronic libraries; to help train others; to hold periodic meetings; and to contribute to all the many other tasks that will be required if the organization is to achieve its vision. This means that "buy-in" is needed and that simply copying the Cochrane Collaboration approach to governance and work processes and applying it wholesale to the EBTC likely will not work.

Kotter (1996) identifies eight steps to leading change. These include establishing a sense of urgency, creating a guiding coalition, developing a change vision, and empowering broad-based action. Kotter also extolls the value of generating short-term wins at the outset of the process.

Fowle consulted a number of sources of information on the establishment of governance, policy, and procedures for a non-profit organization such as the EBTC. "The Basics of Forming a Nonprofit" checklist from "Nonprofit Law and Governance for

¹¹ <http://www.cochrane.org/>



Dummies” (Welytok, 2007) encapsulates advice from various publications:

To ensure the success of your nonprofit organization, you need to start with a solid foundation. Take a look at the following fundamentals checklist so your nonprofit is set up properly and legal issues are covered right from the beginning.

- Clearly define your mission and its scope
- Put together a business plan and system
- Adopt a set of bylaws
- Recruit a board
- Hold an organizational meeting and define duties and responsibilities
- File for tax-exempt status with the IRS¹²
- Register with your state
- Get staff and volunteers in place

It is of critical importance to clearly define the mission and scope of any organization. Beginning with the end in mind (Covey, 1989) allows one to visualize what success will look like in the future by providing a “compass” to help direct the actions toward the ultimate goal. Thus, the EBTC’s mission and vision are key elements to consider when establishing procedures for governance and work processes. They define what job the organization is to do and what a successful future would look like. All that the EBTC does will derive from and/or be evaluated against its mission statement.

In addition to the purpose and nature of the organization, a mission statement has implications for what the organization will do and how it will do it. A “to do” list can be developed from a properly crafted mission statement to guide the operations and governance of an organization.

Although the EBTC has not yet adopted a mission statement, it does have an early draft of one, which is admittedly overly long: *The EBTC will facilitate the systematic, objective, and transparent assessment of test methods to foster the increasing use of in vitro test data in making safety decisions for human exposure. The EBTC will facilitate the adaptation/development and use of systematic reviews of methods to identify the most promising existing/emerging methods so that these may be used more broadly to generate critical safety data, [as well as] tools for data appraisal, meta-analyses, and test assessment methodologies, to identify the most promising existing/emerging methods so that these may be used more broadly to generate critical safety data. The EBTC also will apply these tools and approaches to the construction of testing strategies to ensure effective and efficient testing, as well as to the synthesis of data across studies to facilitate robust decision making.*

This draft mission statement implies that the EBTC wants to add value to 21st century toxicology, as well as to traditional toxicology, by speeding the adoption of the best methods for next-generation safety assessment testing. It hopes to do this by adopting the tools and approaches of EBM as used by the Cochrane Collaboration. A primary activity implied by the draft statement is to sort out the difficult issues associated with

reviewing data on chemical toxicity from multiple types of methods and integrating these data to inform decisions about chemical safety. In other words, how do you integrate data from animals, tissue culture, high-throughput tests, various “omic” technologies, *in silico* approaches, etc. when the results do not all point in the same direction? A final task implied by the draft mission statement is to apply evidence-based thinking to the various existing and new toxicity testing approaches, perhaps through an adverse outcome pathway approach, to build an efficient integrated testing strategy or to critique a proposed strategy in an unbiased fashion.

A well thought-out and carefully crafted vision statement is just as important as a mission statement to an organization. While a mission statement speaks to the “head,” a vision statement speaks to the “heart” by providing the members of the organization with a picture of how things could be in the future and how that is of value to them as individuals. The vision is a stretch goal to strive for. Effective vision statements are easily grasped by all, but are very difficult if not impossible to reach, because they are designed to stimulate continual improvement as well as to provide a sense of purpose and a common goal.

To aid the EBTC leadership as they develop a vision for the organization, Fowle posed the following questions, derived from the Cochrane Collaboration’s retrospective analysis¹³:

- What is the organization (what are our aspirations for it)?
- Who does it serve (who are our customers)?
- Who wants what, when, how (what are the needs and expectations of our stakeholders)?
- What will the organization provide (what are our products and services and what are the needs of third party users who will leverage our products and services into their own)?
- How are we differentiated from others?
- What is in scope?
- What is out of scope?

A similar set of questions also could help the organization fine tune its mission statement, he pointed out.

Fowle also talked about the value of having a solid business plan and pointed workshop attendees to sources for advice on writing a business plan for a nonprofit corporation¹⁴.

Fowle acknowledged that properly developing and adopting a set of bylaws may seem too tedious and time consuming for many, given the enthusiasm to get on with advancing a new organization’s mission. However, in the spirit of “going slow to go fast,” he argues that it seems worthwhile to take whatever time is needed to place the organization on a firm footing to avoid misunderstandings, wasted time, and potential hard feelings in the future. He pointed out that “Robert’s Rules” comprehensively classifies organizational rules based on their application and use and on how difficult they are to change or suspend (Jennings, 2006). He recommended that the EBTC consider adopting Robert’s Rules of order, amending them as needed to meet the organization’s needs.

¹² Internal Revenue Service of the US federal government.

¹³ <http://ccreview.wikispaces.com/file/view/Collaboration+review+-+Recommendations+Report+-+FINAL+2009.pdf>

¹⁴ <http://smallbusiness.chron.com/write-business-plan-nonprofit-corporation-3061.html>



The EBTC also will need to file for tax-exempt status with the IRS and register with the state, and Fowle presented sources for doing these things, as well as for hiring staff.^{15, 16, 17}

The Cochrane Collaboration's work processes and governance are described in the "Newcomer's Guide" on the Cochrane Collaboration website.¹⁸ Further details about governance and work processes can be found in The Cochrane Manual.¹⁹ Cochrane entities receive their funding from different sources but agree to follow the policies and practices of the Cochrane Collaboration.

The Cochrane Collaboration also employs two ombudsmen to help resolve areas of conflict that arise between people or entities, for which the usual process of involving their Centre Director has not been sufficient.

The Cochrane Collaboration's central functions are funded by royalties from its publishers, John Wiley and Sons Limited, which come from sales of subscriptions to the Cochrane Library containing the publications produced by the Cochrane Collaboration. The individual entities of the Cochrane Collaboration are funded by a large variety of governmental, institutional, and private funding sources and are bound by organization-wide policy limiting uses of funds from corporate sponsors. The Cochrane Collaboration's sources of support can be found online.²⁰

Fowle offered his thoughts on how well the Cochrane Collaboration model meets the EBTC's needs. While the EBTC has methodological and informational components like the Cochrane Collaboration efforts, the audiences of the two entities are likely to be different in a few important aspects. The medical orientation of the Cochrane Collaboration is likely more accessible to individuals than is toxicity test qualification, which is a central focus of the EBTC, given that individuals are directly engaged in health care issues but are not directly engaged in the types of decision making that occurs in industry and regulatory agencies. Additionally, the various health outcomes that the Cochrane Collaboration deals with are real and can be seen in patients by physicians daily, while toxicity outcomes in human populations are sometimes theoretical because of measures taken to avoid any adverse outcomes. Finally, the Cochrane Collaboration deals with human health issues only.

Because toxicology deals with environmental health as well as human health, Fowle suggested that the EBTC may wish to adopt this two-prong focus so that it is able to address the environmental health mandates of US agencies such as EPA, the National Marine Fisheries Service, the National Wildlife Service, the United States Geological Service, and corresponding state entities that are responsible for assessing risks to wildlife and ecological systems.

Fowle raised points to consider for a future discussion about the formation, governance, and work processes of the EBTC. The EBTC's success will depend in large part on the enthusiasm and engagement of many people of different backgrounds and

in different locations to do the work of the organization. He suggested that it would be possible to use questions employed by the Cochrane Collaboration's retrospective analysis (see Section 6.2) to elicit input from the interested stakeholders about what the EBTC should be, what it should look like, and what it should do to maximize the chance of success.

Finally, Fowle offered two proposals aimed at triggering discussion and identifying action items and next steps:

1. Establish the EBTC as a nonprofit organization following the general approach outlined in this white paper.
2. Base the EBTC Governance and Work Processes on the specific model of the Cochrane Collaboration, including its principles, with modifications as appropriate to suit the EBTC needs.

6.2 Invited discussant Roberta Scherer

In response to the points John Fowle raised about the steps needed to create an organization and procedures for the EBTC, Roberta Scherer of the Cochrane Collaboration and Johns Hopkins University described how the Cochrane Collaboration has addressed these issues over the years.

She began by describing the evidence-based health care process. Cochrane becomes involved after evidence on a particular medical topic – often in the form of a clinical trial – is generated. This evidence may or not have been published. Cochrane conducts a systematic review bringing the evidence together, synthesizing it, and doing a meta-analysis on it. This evidence is then used by professional societies or others to make decisions. It also may lead to policy that is applied in the healthcare setting.

Scherer underscored that the evidence is only one aspect of the situation. The other aspects – what are used to apply the evidence – are clinician expertise and patient values.

Scherer noted that the collaboration's name honors Sir Archie Cochrane, a British epidemiologist. In 1979, he criticized his profession for failing to "organize a critical summary by specialty or subspecialty adapted periodically of all relevant randomized control trials." In the 1980s, Sir Ian Chalmers, an obstetrician, took up Cochrane's challenge and organized a group of people to develop the Oxford database of perinatal trials. The group gathered all of the randomized trials having to do with perinatal care.

In 1992, the British National Health Service funded the first Cochrane Center. The Oxford database of perinatal trials that Chalmers amassed in the 1980s served as a proof of concept study and precipitated the idea that systematic reviews of randomized controlled trials testing various medical interventions could be done for all of medicine. The purpose of the original center was to facilitate the preparation of these systematic reviews. The Cochrane Collaboration as such (not just a center at Oxford) was launched in 1993.

¹⁵ <http://www.sos.state.md.us/charity/non-profit.aspx>

¹⁶ <http://www.sos.state.md.us/charity/checklistnonprofit.pdf>

¹⁷ <http://www.answers.com/topic/nonprofit-organizations-and-human-resources-management>

¹⁸ <http://www.cochrane.org/about-us/newcomers-guide>

¹⁹ <http://www.cochrane.org/policy-manual/welcome>

²⁰ <http://www.cochrane.org/about-us/funding-support>



The Cochrane Collaboration presents summary data in the form of “forest plots” to graphically depict the trials used in evaluating evidence on a particular topic and the level of concordance between the trials, Scherer told attendees. The plot includes a horizontal line for each trial evaluated; the line’s length is based on its 95% confidence limits for the piece of evidence in question. A vertical line indicates the area of no difference, and a small diamond denotes the pooled result of the meta-analysis. The width of the diamond represents the confidence in the result, in terms of the meta-analysis.

The Cochrane Collaboration’s “product” is *The Cochrane Library*, which Scherer said is widely considered to be the single best source for reliable information on the effects of healthcare. The Cochrane Library includes the Cochrane Database of Systematic Reviews. The database includes more than 4,500 full-protocol reviews, as well as the published protocols themselves and methodology reviews.

“We publish our protocols before we do the review; the protocols are also peer-reviewed by experts to make sure that our reviews will be relevant and important,” she said. The library includes a register of all of the clinical trials that have been identified via electronic and hand searching.

The methodology reviews are systematic reviews of EBM/HC research methods. The Cochrane methodology register is a collection of studies pertaining to research methodology.

Scherer went on to describe a Cochrane review as a review of existing knowledge that uses explicit and scientific methods. All reviews are crafted at the outset to include a clear description of how they are going to be done, and all are designed to follow the rules the organization has created to ensure that all of the information about the review process is transparent to all observers.

Cochrane reviews include a clear description of the research question. To avoid duplication of effort, each review is registered before the protocol is written to ensure that the questions are relevant, important, and have not already been asked by someone else. Each review specifies the inclusion and exclusion criteria for the randomized control trials that will serve as evidence. The review stipulates how the reviewers are going to identify this evidence, including which databases they will search, and whether any searching will be conducted manually. Reviewers define the search strategy they will use before they begin because that will be part of the publication. Reviewers specify the methods that they will use to assess the risk of bias, as well as the methods that will be used both to extract and to summarize the data. Studies are examined closely to see if they are similar enough to pool study results, i.e., whether there is clinical heterogeneity. Excessive clinical or statistical heterogeneity would preclude pooling study results.

Scherer averred that the strength of Cochrane reviews is their rigorous methodology. They are collaborative efforts, and it is not unusual to have an international group of authors.

Scherer also pointed out that reviews are performed independently of industry funding to avoid the perception of bias.

Scherer then provided details about the Cochrane Collaboration on five points to aid in comparing it to EBT.

1. **Mission.** The Cochrane Collaboration is international, and it includes people in both developed and developing countries.

The collaboration translates Cochrane reviews into other languages.

2. **Values.** The Cochrane’s values are the foundation for what the organization does. They explain how the organization achieves its mission. The organization’s values include collaborating, avoiding duplication, minimizing bias, keeping up-to-date, ensuring relevance, and ensuring access.

3. **Business plan.** The Cochrane Collaboration exists for people who need the evidence – healthcare providers, decision-makers, consumers, and researchers.

4. **Funding plan.** The initial funding for Cochrane was by government agencies, especially those in the U.K. The organization became a nonprofit in 1996. Royalties from the Cochrane library are used for Collaboration-wide endeavors and to support annual meeting costs. Other than that, each entity is required to obtain its own funding.

5. **Bylaws.** Cochrane developed a policy manual that is available online and updated as needed.

The first thing the Cochrane Collaboration did was to hold an organizational meeting. It included 70 people from nine countries. The organization now has 28,000 people in more than 100 countries, Scherer said.

The Cochrane Collaboration steering group’s first meeting was in 1994. Scherer said that the steering group focused on incorporating the organization and planning the bylaws, as well as creating the software and discussing how to create a handbook to describe how to do the reviews.

Soon after the organization was launched, it developed software for conducting reviews. The organization also developed ARCHIE, an online repository for directories, reviews, and files shared by people within the Collaboration – similar to current cloud computing. The organization also has developed training materials and sponsors workshops to train people in conducting systematic reviews.

The Cochrane Library was created very soon after the formation of the organization. It was initially published and made available online in 1996, and the online Cochrane journal is now published monthly. More recent activities include the review of diagnostic tests.

Scherer concluded by exhorting the audience not to forget about what she called “the heart.” The operation of the Collaboration has not always been a smooth ride. However, because everyone involved in the Collaboration shares the same vision – the heart – it has been possible to get through tough times. She said that while the EBTC needs to come up with bylaws, it is necessary to be flexible – what the EBTC chooses to do now may not be relevant in 15 years. Finally, she said that the transparency and methodological rigor based on the empirical studies that underpin the Cochrane reviews are worth emulating.

6.3 Invited discussant Rashid Shaikh

Rashid Shaikh of the Health Effects Institute (HEI) described a different organizational model for the EBTC, one that has proven successful in the environmental health and toxicology space. Rather than a detailed description of how to set up the organization, he focused on describing what the HEI is, how it is organized, and what it has achieved.



He stressed that establishing organizations as independent and impartial is critical to establishing credibility with both the stakeholders and the larger public. Independence and impartiality also are intertwined with the other key issues of funding and governing structure. Shaikh emphasized that communication with the scientific community, the public, and decision makers, is also very important.

Shaikh explained that HEI was formed in 1980 in the context of the Clean Air Act's section §202(a)(4), which required automakers to test the health effects of automotive emissions. At that time, the science on which decisions about air quality standards were based was highly contentious. Industry leaders and EPA officials were motivated to find a better way to produce science that all parties would find trustworthy.

According to Shaikh, HEI is structured to maintain credibility and transparency on scientific issues pertinent to regulatory questions in the air pollution and health areas, which are often controversial. The organization strives to maintain a balance of government and industry funding. While the EPA and the motor vehicle industry provide a significant portion of HEI support, other government and private groups also support the organization.

HEI has an independent board of directors and two scientific committees, Shaikh explained. The people serving in these capacities are respected senior leaders, recognized for integrity and scientific accomplishments, and are not employed by the auto industry or other government agencies associated with Clean Air Act regulations. He added that HEI does not take policy positions and that its work is solely focused on scientific research and evaluation.

HEI's mission is to provide independent, impartial, high-quality and timely science on the health effects of air pollution. The outlines of HEI's work are planned in advance through a comprehensive process: HEI consults with EPA and automotive industry sponsors, as well as other agencies, including the National Institute of Environmental Health Sciences (NIEHS), the Department of Energy (DOE), and the Federal Highway Administration (FHWA). The scientific community and a wide variety of NGOs and other groups are also invited to offer suggestions. Through this process, HEI develops a blueprint for five years of work, which is termed their Strategic Plan.

Once the Plan has been completed and the outlines of the research agenda have been set, the research to implement it begins. HEI develops requests for applications and funds research through a competitive process. Research is overseen by HEI staff and the Research Committee. At the end of the research phase, each investigator prepares a report, which describes the work in far greater detail than scientific papers. Next, the Review Committee, which has no role in selection or oversight of the research, conducts a peer-review of the report and prepares a commentary. HEI publishes the report and commentary on its website and in print.

From time to time, HEI also prepares broad reviews of a field of science, for example, health effects of exposure to traffic-related air pollution or health effects of ultrafine particles.

HEI selects its board of directors after consultation with HEI's core sponsors, including the EPA Administrator and a majority of the motor vehicle industry. The scientific committees

– including the Research and Review Committees and special review panels – are appointed by the Board; for major appointments, such as chairs of committees, HEI sponsors may also be consulted. The board, scientific committees, and staff all must undergo a conflict of interest disclosure process annually.

Shaikh emphasized that one of HEI's key attributes is transparency, which includes full public disclosure of all results, both positive and negative. HEI reports are detailed and comprehensive, and the accompanying commentary provides both the context and evaluation of the work.

Shaikh stressed that HEI also is committed to allowing access to data and other details to the entire scientific community. Many HEI reports include extensive appendices that have original data or detailed data summaries. Following publication, access to data and methods are made available to other investigators.

Other partners have joined HEI over the years, including other government agencies, the oil industry, and some chemical companies. Other organizations that support HEI's work include the DOE, the FHWA, the American Petroleum Institute, CON-CAWE, Hewlett Foundation, the Asian Development Bank, and others.

6.4 Invited discussant Dennis Devlin

Dennis Devlin, ExxonMobil's environmental health advisor, is also a board member and current president of the Health and Environmental Sciences Institute (HESI), whose mission is to bring scientists from academia, government, and industry together to address health and environmental issues. He was representing HESI as a discussant.

Devlin complimented Fowle on his white paper, which he said served to lay down the foundation for furthering EBT. His goal, he explained, was to discuss practical considerations related to putting together a nonprofit organization and to respond to the questions that Fowle suggested he address.

Devlin felt that the mission and vision were key statements for defining an organization, ensuring long-term commitment, and reaching out to potential funders. He pointed out that a mission should succinctly describe why the organization exists and that the vision should describe the organization's aspirations for what it will achieve. For both a mission and vision, Devlin stressed the importance of being precise and transparent so people have no doubt about why the organization exists. In considering the draft EBTC mission statement, he opined that it read more like a mixture of a mission statement, a vision, and work processes. All of these are very important, but each separate statement has a concrete purpose.

Devlin proposed a few mission statements for the EBTC for attendees to consider, such as "to add value to 21st century toxicology by speeding the adoption of the best methods for next-generation safety assessment testing." Next, Devlin proposed a vision statement for the organization: "Enhanced decision-making for safety assessments based on the development, assessment, and adoption of evidence-based methods and testing strategies."

Devlin said that it was important for a science-oriented nonprofit organization to divide its governance and procedural is-



sues into two camps – administration and the scientific process of the output – and to consider these issues separately. Non-profit organizations require a number of administrative tasks to be completed, including developing a business plan and a board of directors, which would be a group of experts separate from the scientific Steering Committee. It would not be appropriate to put the top-notch scientists on a board of directors dealing with budgets.

The administrative plan requires external expertise, including a facilitator, lawyers, finance, communication, and human resources. Devlin felt that the plan proposed by Fowle represented a good starting point in incorporating *Nonprofit Law & Governance For Dummies*.

Devlin noted that developing a strategic scientific plan will take time and a diverse group of dedicated stakeholders, but the effort is certainly worth it. One of the steps that such a group would take would be to confirm and adapt the mission and the vision and use them to develop the strategic plan.

To illustrate how strategic planning can work, Devlin explained how HESI does it. HESI's scientific plan is always five years out. To meet that challenge, HESI has four strategic priorities. For each priority, there are supporting objectives. Importantly, there is a background document that clearly articulates accountability. Devlin showed the group HESI's strategic plan for 2011-2015 to illustrate how this kind of information can be presented concisely and clearly to keep everyone on the same page.

Devlin remarked that the EBTC's mission statement should be crafted in such a way that everything the organization does and/or could do can be evaluated against it.

In response to the question of "what are the needs and expectations of our stakeholders," Devlin said that he didn't know the funding mechanism and didn't see relevant details in the documents. He asked whether the funders have special interests. For an organization like this, having short-term success is important both to maintain morale and to ensure funding, he said. He advised the EBTC to try for a success, even a small one, within 3-4 years.

Devlin suggested that the organization should provide tools to assess methods in the "increasingly diverse toxicology toolbox," as well as reviews identifying the most promising existing and emerging methods. The EBTC may maintain databases and an electronic library to provide access to guidance and assessment documents.

An important issue is how the EBTC is differentiated from other groups and that, in order to do this, one must consider who the "others" are. Devlin believed this particular endeavor has many "others" and therefore the EBTC needs to differentiate itself and then to actively decide if it will collaborate or compete for both experts and dollars. The presumption is that EBTC will be collaborating, because working together and dividing the pieces can be a good way to produce useful results.

In terms of structure, Devlin recommended simplifying the Cochrane Collaboration's approach, which he thinks has merit. He observed that review groups that find "evidence" and pre-

pare, maintain, and update reviews may make sense, for example. The question is: Will the EBTC provide the information for free? (The Cochrane reviews are available via subscription.) Review groups should likely be formed around certain technical areas, beginning with something that is achievable, perhaps developmental toxicity.

Devlin said he also felt that methods groups would be useful to improve methods to capture and use evidence, conduct, and apply reviews. Some kind of a users' network would be useful, he noted.

The board and the scientific steering group, as well as the funders, will decide what structures/entities should be established. The board also will determine who the leaders are. In the US, nonprofit organizations require boards of trustees to be responsible for the oversight of the entire enterprise, including the chair and vice chair, secretary, and treasurer.

The role of the executive director is critical, Devlin said. That person must focus on short- and long-term strategies, staffing, and many other important issues.

The steering committee is going to focus on the technical content. It could be a subcommittee of the board, he pointed out. Membership for the board and steering committee could begin with the existing steering committee. Key questions to consider include whether funders require representation and whether there is a future role for official members/participants. He said he doubts whether EBTC needs all positions of the Cochrane Collaboration. Fit for purpose is what is needed, he said.

Devlin thought that one area where the Cochrane Collaboration may not be a good starting point is in developing the EBTC's bylaws, partly because Cochrane is essentially a UK-based organization. It would probably be better to work with a group more familiar with US law. Devlin concluded by showing the bylaws for HESI, which he said were quite a bit simpler than the Cochrane Collaboration's bylaws.

6.5 Invited discussant Andrew Rowan

The final discussant was Andrew Rowan, who is the CEO of the nonprofit Humane Society International, as well as a member of the Human Toxicology Project Consortium, which seeks to accelerate pathway-based approaches to toxicology.²¹

Rowan welcomed the attention to good management practices, arguing that not nearly enough attention is paid to such issues in the nonprofit world.

Rowan counseled that the name "evidence-based toxicology" may be off-putting to some, given that "all of us – I would hope – are using evidence."

He questioned the frequent calls for impartial people in endeavors such as the EBTC. "There's not an individual in this room who is impartial, in the sense of being completely blank without opinions, biases, or prejudices," he contended. "I would much rather that we have partial people who are identified with particular positions arguing vigorously on behalf of their positions," he said.

Part of the impetus for the Cochrane Collaboration was that peer-review was not doing what it should be doing, Rowan said.

²¹ <http://htpconsortium.wordpress.com/>



Peer-reviewed papers are no doubt preferable to papers that have not gone through peer-review, but they are not necessarily great, he said.

In the context of the need to objectively assemble evidence, Rowan argued that emotion is an important part of decision making. He cited the work of Antonio Damasio, such as *Descartes' Error*, in revealing how important emotion is in decision making. Damasio observed that when emotion is eliminated via a frontal lobotomy, people's ability to make decisions is impaired. Lobotomized people cannot distinguish between important and unimportant decisions, Rowan explained.

"In fact, the human brain is a rather poor mechanism for deciphering large sets of conflicting variables," Rowan said. We tend to eliminate a whole lot of stuff because of emotional commitments that have been built up over the years, and then we can make a decision between two or three options, he observed.

People are also good at intuitive decision making. Efforts to program computers to be chess masters have revealed how complicated intuitive decisions really are, he said.

Rowan's plea to the group was not to seek complete impartiality by avoiding corporate funding. "Corporate funding can be very useful," he observed. "It can bias results if you let it, but if you have NGO activists in the system, it probably won't." He agreed with his predecessors on the value of transparency.

Rowan agreed with some researchers who have argued that most animal research is poorly designed. Much of it is not double-blind, for example, he said. There are interesting issues involved in what we accept as evidence and how much we will accept as evidence, Rowan continued.

Rowan expressed his belief that mission and vision statements should be as concise as possible, as well as memorable. He then proposed for the EBTC mission statement, "to enhance trust in and integrity of toxicology and regulatory science" and for the vision statement: "to create a safer world for humans, animals, and the environment." He posited that everyone in the room could agree that these were worthy goals.

Rowan argued against following the Cochrane Collaboration's path into becoming a publishing company. "Having to pay a publishing company for access to these materials, to me, is not transparent," he said. "I think transparency requires open access."

Funding issues are important, Rowan continued. He urged the EBTC to be completely transparent regarding where its money comes from and where it goes. Although he said that this information is available for the Cochrane Collaboration in the U.K., he was unable to find funding information for its U.S. arm.

Rowan concluded by reiterating his observations that corporate money might prove helpful to the EBTC and that transparency is likely to serve the organization well.

6.6 Open discussion

Ellen Silbergeld opened the discussion by commenting that a good next step might be to try out some examples in an effort to find out how to operationalize the concepts that had been discussed during the workshop. She said that this could build confidence and shape the way that the vision statement is formulated. There are still a lot more things that need to be thought through

about how this will actually relate to and improve whatever our separate visions are of toxicology.

She proposed that the next step might be to put out an open call for nominations for crisp proposals that might fit into two areas:

1. Applying EBT tools in evaluating a method. For example, all *in vitro* tests for estrogenic activity might be evaluated for evidence of disruption of estrogenic function in humans.
2. Looking through the toxicological literature on dioxin and see if one can do a systematic review examining the evidence associating exposures to dioxin with developmental toxicology – or some other endpoint that one could succinctly phrase.

Silbergeld said that she would be wary of joining an organization aimed at promoting something like EBT without having a sense that this really is something that our field is ready to incorporate. She also argued for a broader vision beyond advancing *in vitro* methods. Think of Cochrane's vision, which is to improve healthcare, she pointed out. She therefore suggested that although one component of the EBTC's mission could be to advance new methods, restricting it to that would limit consumer and public interest in and support for this activity.

John Fowle asked Roberta Scherer to say more about the 1993 pilot project that led to the creation of the Cochrane Centers and how it served as a driver that gave the organization momentum. Scherer said that the initial development of the Oxford Database of Perinatal Trials was a proof of concept that showed that evidence-based medicine evaluations were possible. However, she also noted that a concerted effort to do systematic reviews was happening at the same time. "It was an intersection of a need and the methodology being there and developed at the same time," she observed. The Oxford Database of Perinatal Trials showed that this intersection could result in something that was valuable, she said.

Fowle said that his research drove home the point that it takes a lot of work to set up an organization and, consequently, it is necessary to be fully committed to it and have clear objectives for it. To that end, he said it could make sense to develop a confidence-building pilot project first. Shaikh agreed, suggesting that a pilot activity could be undertaken while some other things are being put into place.

Naidenko observed that it is important to consider who the organization's users or clients will be. The Cochrane example seems like it was a lot easier to pull off.

Fowle said that a pilot project could generate an "early win" and thereby help drive funding and support for EBT. He agreed with Naidenko that considering who the organization's clients would be may also be helpful in this regard.

Terry Quill of the Quill Law Group, who was trained as a toxicologist, observed that the conference seemed to play out something like a roller coaster ride. He concluded that EBT for the 21st century is associated with two issues. The first is evidentiary science and how toxicology should deal with that in the future with the idea that something different is needed. The second issue is the new 21st century techniques and how the field is going to deal with them. "I'm afraid that what's happening is that, depending on who is talking, there's an emphasis on one side or the other."



Quill asked the participants if what the group wanted to do was to use weight of evidence in toxicology in ways it has not yet been applied. “If that’s what this group wants to do, there’s certainly a need for it,” he commented.

Quill observed that the new techniques now available to toxicologists have many potential benefits, including saving money and time, as well as facilitating better health decisions. “On the other hand, we could spend more money if we have higher false positive rates,” he said. To him, this raised the question of what the group was going to do to see that the techniques are properly validated, relevant, and useful.

Validation is going to be key, Quill continued. “I’m afraid of the possibility that we may just throw these assays out there and people will misuse them and misuse the data,” he said. If that happens, he posits that the field will be in a worse position than it is now.

Rowan commented that the key difference between *in vivo*-based toxicology and *in vitro* systems is the volume of data the latter produces. Within a year, we could test all the 30,000 REACH chemicals for which there is currently very little data and produce some biological data with 200 different assays at 15 different concentrations. It’d cost about \$ 1500/chemical,” he said.

“What the data means is of course another issue,” Rowan continued. But he argued that the whole-animal assays conventionally used in toxicology studies have never been validated, beyond the fact that they were being conducted on mammals and humans are mammals.

According to Rowan, “[w]hat we’re doing currently is not going to get us very far very fast. Even if we have more questions, we’re not going to do more animal testing because the capacity isn’t there to do it. The answers also aren’t there from historical animal testing. Do you think \$ 35 million for more BPA studies is going to produce a clearer outcome for BPA? I don’t. Probably what will happen is the corporations will slowly move away from it because it’s problematic.”

“It’s the volume of data that’s going to be the real driver of this, and the fact that we’ll have to interpret it as we go along,” Rowan summarized.

Richard Woychik of NIEHS felt that Silbergeld’s comments at the beginning of the open discussion were spot on. “It’s very important to define what EBT is and what it’s not. I hope that it’s not just *in vitro* studies. While at some point way down the road a series of *in vitro* studies may be predictive for toxicology, I don’t think we’re there yet. Nor do I think we’re going to be there in the near term.”

Finally, Woychik said that he agreed with Silbergeld about the need for a crisply defined pilot study with some specific objectives.

Thomas Hartung thanked the people who contributed to the discussion for their thorough analysis and the excellent food for thought that it provided. “It’s most important that we keep the ball rolling now, because we are at a critical stage of forming something that is more than coming together.”

The transparency aspect and financial considerations have been mentioned several times, he pointed out, and he said that he wanted to disclose some of the details the he had not previously mentioned.

We at Johns Hopkins are very much committed to EBT, Hartung stressed. His endowed chair is in the denomination of EBT because the Doerenkamp-Zbinden Foundation in Switzerland, which endowed the chair, and Johns Hopkins agreed that this is a worthwhile effort. This is a continuation of the work that Hartung began at ECVAM.

That said, Hartung stressed that his institution’s commitment to EBT does not imply any ownership. “We don’t want to own the process, but instead to help it along, and we will be very happy if the process is taken further by others in the future,” he explained. EBT is not intended to be focused on alternative methods under a new name, he stated. Nor is it toxicology of the 21st century under a new name. It is meant to be something that adds a component that is lacking in the field of toxicology, he said. It is a driving force towards guiding the way we handle information. It is not meant to produce new methods or to replace methods. It is meant as a process of bringing a certain type of thinking into our debates, which might be beneficial for all of us.

Johns Hopkins’ EBT effort has continuing support from the Doerenkamp-Zbinden Foundation. This will enable it to provide very limited funding for carrying out workshops and dedicated small studies via organizations such as the Transatlantic Think Tank for Toxicology (t⁴). We are trying to provide compensation to individuals who want to conduct some type of systematic reviews. At present, the amount available for agreed, dedicated studies is € 10,000 each, which at the time of the workshop equated to between \$ 13,000 and \$ 14,000. Hartung said he hoped that the funding would help motivate people to initiate reviews and to compensate them for some of their costs. It will be linked to the opportunity to publish the review in ALTEX, which currently has an impact factor of 4.4, he said.

Hartung said he is discussing with his Center’s management the possibility of providing additional funding for encouraging the production of systematic reviews in the future. He said he had between \$ 200,000 and \$ 250,000 per year to sponsor research, and some of this might become available for systematic reviews.

Hartung said that a philanthropic donor has provided funding to establish and maintain the EBTC secretariat; this represents the most important source of funding via an agreement that is valid for five years. The funding supports Martin Stephens as the organization’s director, as well as a support person. It also includes funds for supporting steering group travel. Taken together, this suggests that the EBTC has some foundation for future achievements. The next steps are to move ahead to discuss the organization’s structure, as well as potential pilot projects. He concluded by saying that the day’s discussions had been very helpful.

7 Discussion

7.1 The nature and scope of evidence-based approaches

The workshop served to clarify the nature and scope of evidence-based approaches in general and EBT in particular among the participants. Some participants equated EBT with test methods validation, rather than as a set of approaches and tools to as-



sess the evidence on any well-framed question in toxicology, only *inter alia* including questions regarding test method performance. By the close of the workshop, participants confirmed that they better understood evidence-based approaches and the applicability of EBT. Participants also grasped the distinction between EBM/EBT's distillation of the evidence and any policy decisions that others might draw from the evidence.

This greater understanding of evidence-based approaches was largely a consequence of the presentations by Ellen Silbergeld (Silbergeld and Scherer, 2013) and Roberta Scherer (Section 6.2), as well as their comments and Thomas Hartung's comments during the open discussions. The nature and components of a systematic review became clear. In addition, the following points were emphasized:

- Any policy-related decisions informed by systematic reviews are based not only on the evidence, but also on professional expertise and – in the case of EBM/HC – patients' values.
- An organization's values are an important part of its identity, and the resulting shared vision can carry an organization through difficult times.
- Taking industry funding for particular reviews could be perceived as a source of bias.

In addition to a general orientation to evidence-based methods in medicine/health care and toxicology, Ellen Silbergeld mentioned examples of some of the few systematic reviews that had been conducted in toxicology, an approach that she and her colleagues have pioneered. These case studies examined the effects of lead and arsenic on human health and illustrated the use of forest plots (Navas-Acien et al., 2006, 2007). These were examples of the application of evidence-based methods to hazard assessment. The written version of Silbergeld's presentation, co-authored with Roberta Scherer (Silbergeld and Scherer, 2013), takes a somewhat more conceptual approach to the subject, providing a practitioners' view of the nature and importance of evidence-based approaches in medicine/health care, as well as the potential and the challenge of translating these approaches to toxicology.

The commentary by Kim Boekelheide – who is by training a physician familiar with EBM/HC – noted important differences between clinical medicine and toxicology, including the diverse experimental approaches in toxicology, which would make evidence synthesis more challenging (Section 4.2). He raised the important question of the extent to which EBT approaches might be able to prospectively guide new studies in a field, in addition to providing retrospective analyses of past studies. Others noted that once an EBT assessment has been completed, it also can provide feedback and inform how new data is generated, the types of studies that are conducted, and how studies are written up for publication. This feedback loop can help reduce the elements of bias and increase study quality in the discipline as a whole.

George Woodall of US EPA's National Center for Environmental Assessment noted that many of the recommendations made during the workshop are also made in the NRC's recent review of the IRIS program's assessment of formaldehyde (NRC, 2011). He noted that those considerations are being integrated into the IRIS program, which was encouraging to EBTC members.

7.2 General themes and recommendations

Several general themes and recommendations emerged from the workshop. First, the learning process during the workshop (see Section 7.1) led many participants to suggest a primary role for the EBTC in education and outreach on evidence-based approaches.

A second over-arching theme was an obvious need to demonstrate the value of evidence-based approaches to toxicology through case studies. Two of the three presentations featured as "case studies" in workshop Session 2 (Daland Juberg's and Patricia Harlow's – see Keller et al., (2012) and Section 4.1, respectively) have clear relevance to *future* work applying evidence-based approaches to 21st century methods, but given the current embryonic state of EBT, these presentations were necessarily not demonstrations of how EBT had already been applied. However, Silbergeld's presentation in Session 2 (Silbergeld and Scherer, 2013) did summarize some evidence-based studies of the literature associating specific chemicals with certain adverse effects. Nonetheless, it was clear that the EBTC should organize a number of case studies to further demonstrate the value of evidence-based approaches to toxicology. Such early "wins" are likely to be important in securing stakeholder buy-in early in the process.

A third theme, recurring throughout the workshop, was the importance and challenge of transparency. Those carrying out EBT reviews should strive to adhere scrupulously to the EBM tenet of transparency. Yet EBT practitioners must cope with limits to transparency in toxicology, such as those imposed by data confidentiality owing to corporate or regulatory policy. Any challenges posed by practices within toxicology would be in addition to those plaguing other fields, such as the tendency not to publish "no-effect" results (publication bias).

Fourth, the workshop repeatedly highlighted the fundamental distinction between advocacy and analysis in the work of an organization promoting evidence-based approaches. The Cochrane Collaboration confines its work in EBM/HC to analysis, leaving advocacy to others who may decide to take up the results of Cochrane assessments and use them to promote changes in healthcare policy. This issue echoes the distinction in toxicology between a test method's assessment versus its regulatory acceptance, with the idea that EBT could address the former but regulators would make the judgment regarding acceptance. Workshop participants seemed comfortable to have the EBTC carry on the assessments-not-advocacy tradition as evidence-based approaches are translated from medicine to toxicology.

Finally, the workshop raised the important issue of who would be interested in the output of the EBTC, i.e., who would be the EBTC's "customers." The main end users of EBT reviews and guidance will be diverse, including:

- Scientists interested in carrying out evidence-based reviews, appraising the quality of published studies that they read, designing their own studies to high standards, and writing up their own studies such that other scientists can better understand and replicate the details.
- Decision-makers in regulatory agencies responsible for approval of new chemicals or re-evaluation of existing chemicals, approval of new test methods, or the formulation of new testing programs



- Decision makers in industry responsible for the safety of new chemicals being developed or existing chemicals already on the market
- Scientists in government, industry, and academia involved in developing and assessing new test methods
- Stakeholders in the NGO community interested in the intersection of toxicology and public health, environmental health, and new test methods

Modernizing toxicological decision making and practices will be especially relevant to regulators in government and scientists in industry who are seeking to respond adequately to (i) regulatory drivers such as the REACH chemicals program in the EU and the Endocrine Disruptor Screening Program in the US, (ii) societal demands to move away from the current animal-based testing methods, and (iii) the responsibility to test the enormous backlog of poorly tested chemicals already in commerce. In addition, technology developers have a financial stake in seeing that their innovations are assessed appropriately and implemented into practice.

7.3 Toxicological and organizational priorities for the EBTC

Apart from the cross-cutting themes identified above, much of the workshop focused on more specific topics related to the priorities that the EBTC might adopt. These discussions revolved primarily around the issues of validation and governance.

7.3.1 Validation and evidence-based approaches

Much of the workshop program reflected the EBTC's desire to apply evidence-based approaches to assessing the performance of new test methods, especially the pathway-based assays of 21st century toxicology. In this regard, the sessions provided food for thought rather than specific advice.

In Session 1, Richard Judson presented a multi-authored white paper that extensively discusses how the current validation framework might be adapted to the assessment of high-throughput, pathway-based assays (Judson et al., 2013). Validation is an assessment of an assay's (or group of assays') reliability and relevance for a specific purpose. Judson focused on the purpose (or "context of use") of prioritizing chemicals for later follow-up with more definitive, lower throughput testing. In this context, he argued that assessing relevance will be a bigger challenge than assessing reliability, given the nature of robotic, high-throughput systems. The white paper stresses several themes that resonate with evidence-based approaches, including transparency, continuous improvement, and data quality. It also envisions the possibility of using evidence-based approaches as a means of expanding the validation toolbox and thereby offering an alternative to the current costly and time-consuming validation framework, which is more suitable to lower throughput methods.

Respondent Doug Wolf argued that a test method's throughput should be key to determining whether its performance is assessed with the current validation approach or a more EBT-focused approach, with lower throughput assays being assessed via the former and higher throughput assays via the latter (Section 3.1).

Regardless of the validation approach taken, respondent Ed Carney agreed with Judson that validating high-throughput assays for the purpose of prioritizing chemicals for more definitive testing is a good place to start assessing their performance. Similarly, assays with a well understood connection to adverse outcomes, such as estrogen receptor binding, are good candidates to carry forward. Moreover, he noted the unexplored challenge of validating groups of complementary assays, rather than single assays (Section 3.2).

In the workshop presentation by Grace Patlewicz and the accompanying commentary by Patlewicz, Richard Becker, and Ted Simon, these commentators, like Judson et al., mention the possibility of carrying out a new generation of performance evaluations using evidence-based approaches (Section 3.3). They argue, however, for a somewhat different evaluation framework adapted from the Institute of Medicine report on biomarkers, which emphasizes analytical validation, qualification, and utilization. Their preference for an approach that involves qualification echoes Patricia Harlow's description of the U.S. FDA's biomarker qualification program (Section 4.1).

While the speakers in Session 1 alluded to evidence-based approaches being applied to the validation of high-throughput assays, the ensuing open discussion began to sketch what that process might look like in practice (Section 3.4). Its elements would include defining the criteria up front, listing the candidate chemicals, getting the planned procedure peer-reviewed prior to execution, making adjustments, and then carrying out the revised procedure systematically.

Transparency would be key throughout, so the process could not only be understood but also could be replicated by interested parties. Establishing performance standards for a validated assay or set of assays would facilitate the process of continuous improvement.

The overall aim of the process would be to establish the fitness for purpose of an assay by demonstrating its scientific basis. This is a type of relevance that is included, but not emphasized, in the current validation framework. It is a scientific or mechanistic relevance, whereas the current framework focuses on empiric relevance, i.e., correlation to results from reference tests (Hartung, 2010).

The focus on test method validation as a suitable target for EBT also predominated in Session 3, "Setting Priorities for the Evidence-based Toxicology Collaboration." In her presentation, Fitzpatrick noted that the new toxicology calls for new approaches to validation, and that assessing fitness for purpose should be tailored to different purposes depending on the relevant regulatory framework (see Section 5.1). Her presentation (prepared with Abigail Jacobs) parallels Judson's presentation on mapping validation principles to the evaluation of high-throughput assays (Judson et al., 2013). Fitzpatrick's perspective was that of a regulator, addressing the importance of incorporating the emerging methods of 21st century toxicology into regulatory toxicology. She noted the inadequacy of the current validation paradigm to assess the performance of pathway-based assays and offered evidence-based approaches as a transparent and structured way to move forward.



The invited discussants and other commentators from Session 3 agreed with Fitzpatrick and Jacobs on using evidence-based approaches to assess the performance of pathway-based assays as one of the goals of the EBTC (Sections 5.2-5.5). Such assessments would be comparable to EBM's assessments of diagnostic tests and could be viewed as providing a kind of quality assurance to the new toxicology's emerging methods. In keeping with the tenets of evidence-based approaches, these assessments should be carried out in a transparent manner. They also should involve diverse stakeholders. Multi-stakeholder assessments may be liable to reflect biases stemming from differences in values and experiences but, fortunately, evidence-based approaches are ideally suited to handle potential bias.

Workshop Sessions 1 and 3 highlighted the importance of pathway-based testing and the need for new ways of assessing the performance of such testing, with a clear call for mechanistic validation established through evidence-based means. Two of the three presentations in Session 2 explored biological questions that will be central to mechanistic validation of pathway-based methods. Daland Juberg addressed the issue of distinguishing chemical effects that lead to adverse versus adaptive changes in pathway-based assays (Keller et al., 2012). This question is at the heart of the NRC vision of toxicology in the 21st century (NRC, 2007). Similarly, Patricia Harlow addressed the evaluation of proposed biomarkers of biological effects, in the context of the biomarker qualification program of the US FDA's Center for Drug Evaluation and Research (see Section 4.1). Once qualified and publicized, such biomarkers can be pursued within their context of use as indicators of the biological process in question, e.g., nephrotoxicity in rats, to facilitate the drug development process. Assessing the performance of a proposed biomarker is comparable to making a diagnosis – how well does the biomarker in question diagnose or predict the purposed effect. In her presentation, Harlow did not go into the scientific details of the review of particular proposed biomarkers, but she mentioned that the diagnostic performance of urinary biomarkers for nephrotoxicity in rats was evaluated by comparison with currently used nephrotoxicity markers using receiver operating characteristic (ROC) curves – a tool commonly used in EBM/HC assessments in diagnosis and likely to feature prominently in the emerging EBT assessments of test method performance.

The workshop also identified several other issues that will need to be addressed when grappling with the validation of assays (high throughput or otherwise), whether by means of EBT or not. These issues include the following:

- The proper focus of a validation exercise could be an individual assay or a group of complementary assays.
- The EBTC should not be perceived as biased in the outcome of validation exercises, such as by giving the impression of a hidden agenda in wanting to see animal-based methods replaced by *in vitro* methods.
- The EBTC may face the potential challenge of seeking to review the literature on the performance of emerging test methods that are so new that few studies have been published on their performance.

- Similarly much of the data on emerging test methods might be in databases rather than published studies, so this would require new approaches to assessment.

Judson suggested that the EPA EDSP might offer a good starting point for an EBT analysis of pathway-based testing. One could assess how well the proposed “EDSP21” assays can serve to prioritize chemicals for testing in the tier-1 battery.

7.3.2 Governance and work processes

While workshop Sessions 1-3 primarily explored the nature and scope of EBT and the types of toxicological questions that the EBTC might address (especially the issue of assessing test method performance), Session 4 explored how the newly formed EBTC might organize itself and carry out its work to pursue its mission efficiently.

In his presentation, John Fowle provided a comprehensive review of the issues that the EBTC should consider as it contemplates establishing a formal organizational structure and work processes (see Section 6.1). He presented a detailed discussion of the Cochrane Collaboration, which could serve as a model for the EBTC to adopt and to adapt. Fowle's more general advice included the following: that the EBTC should strive to appeal to hearts as well as minds, that it seek to achieve some short-term wins to demonstrate its potential, and that it take the time to get its governance and work processes right from the start (“going slow to go fast”).

The invited discussants were complimentary of Fowle's recommendations. There was general agreement on the issue of looking to the Cochrane Collaboration as a model to draw organizational and administrative guidance from, as appropriate. Roberta Scherer provided additional background information on the Cochrane Collaboration (Section 6.2), whereas Rashid Shaikh (Section 6.3) and Dennis Devlin (Section 6.4) discussed the Health Effects Institute and Health and Environmental Sciences Institute, respectively, as additional organization models that the EBTC could learn from. The final discussant, Andrew Rowan, offered advice on a wide range of organizational issues based on his experience in non-profit management and journal editing (Section 6.5).

While addressing administrative issues may be tedious at times, they are of paramount importance to the future success of the EBTC. As Fowle pointed out, how well the EBTC addresses these issues could mean the difference between success and failure in winning the hearts and minds of the hundreds of volunteers that the EBTC is hoping to recruit to carry its important work forward. He noted that these issues may not seem like priorities in the heady days of embarking on a new initiative, but their importance warrants early attention.

Scherer's presentation included historical remarks that provided a sense of how the Cochrane Collaboration grew and evolved over time. Drawing on this experience should be helpful to the EBTC as it grapples with its administrative priorities while evolving over time and seeking to strike the right balance between subject-matter priorities and administrative ones.

Among the many key points made in the governance session were the following:

- the importance of an organization's mission and vision state-



- ments, of periodic strategic planning, and of distinguishing the organization from similar groups;
- the challenges of administering an international organization;
 - the sensitivity over accepting corporate funding for specific projects;
 - the importance of successful pilot projects to demonstrate value, inspire confidence, motivate staff, and attract members and supporters;
 - the need to clarify what EBT is and is not – If EBT applies to all methods in toxicology, not just *in vitro* or 21st century methods, and to all evidentiary-based questions in toxicology, not just those around validation, then this should be stressed.

Thomas Hartung noted that some funding is available to support the work of the EBTC as it moves forward on both sides of the Atlantic. In addition to supporting the secretariat, this funding could support workshops and small-scale studies.

7.4 Next steps for the EBTC

The US EBTC Steering Committee met one month after its January 2012 workshop to consider the presentations, discussions, and recommendations from the workshop. The Steering Committee is establishing work groups to address the emerging toxicological, methodological, and organizational priorities for EBT, many of which were identified at the workshop. For example, a *methods work group* will adapt the core EBM/HC-tools to the toxicology context. This group also will develop tools suitable for toxicological challenges that may not have close parallels in clinical medicine, such as evidence-based appraisal of data from heterogeneous types of studies (e.g., *in vivo* versus *in vitro*).

Specifically, the methods work group will generate guidance on topics such as information retrieval, data appraisal, evidence synthesis, and test method assessment. The resulting prototype guidance will need to be updated frequently as new insights are gained from their applications.

In parallel, case study groups will explore how the guidance is to be used in practice and how the tools perform by addressing the evaluation of either a test method (or test strategy) or a health effect of a compound, or both scenarios. These studies are expected to demonstrate the feasibility of applying evidence-based approaches in toxicology and to underscore their benefits as compared to standard practices. The components of these evidence-based approaches will be completely transparent, allowing ready reproduction and identification of any bias, e.g., the omission of data in a review. Ultimately, biases in toxicology – a widespread but largely disregarded aspect in the field today – will be made amenable to assessment and exploration.

The EBTC also aims to further the conceptual development of EBT and raise awareness of its principles and approaches. The EBTC will evolve into an umbrella organization facilitating the application of evidence-based approaches to toxicology, comparable to the role of the Cochrane Collaboration in EBM/HC.²²

²² <http://www.cochrane.org>

7.5 Concluding remarks

The EBTC will seek to foster a growing interest in the application of evidence-based approaches in toxicology. The application of these new approaches is expected to strengthen the scientific basis of decision making in toxicology and to improve the transparency of research results, decision making, and the reporting thereof. Better-structured publication of information will facilitate evidence appraisal and synthesis when performing systematic reviews.

With respect to validation, adopting and adapting assessment methodologies from medical diagnostics to toxicological tests is expected to fertilize the development of highly relevant and targeted toxicological testing methods and strategies. The use of test methods based on human biology rather than animal/rodent biology will bridge both the gap and the unquantifiable uncertainty of inter-species differences. It is expected that new technologies (biomarkers and -omics) can be more rapidly introduced as standard methods as evidence-based approaches supplant the time-consuming and cumbersome validation procedures of today. In addition, frequent evidence-based reviews of methods will help identify strengths and weaknesses of methods in practice, providing guidance for future improvements and developments.

Bringing new approaches to the assessment of test method performance will be particularly timely. As toxicology moves to pathway-based approaches (as exemplified by the NRC report on *Toxicity Testing in the 21st Century* (NRC, 2007)), the field will need new tools for assessing test method performance, especially as the focus shifts from animal to human biology. Similarly, as test methods are developed to assess effects at multiple levels of biological organization (e.g., organ on a chip), tools will be needed to synthesize such data in ways that are transparent, objective, and systematic. Ultimately, this effort will open up new approaches to hazard and risk assessment with the ability to flexibly integrate new evidence or adapt to it.

A modernized toxicology allowing transparent, objective, and consistent decision making based on the best and latest available scientific evidence will increase confidence and trust of the stakeholders in government, industry, academia, as well as NGOs who have an interest in and are affected by these decisions.

Modernizing toxicological decision making and practices will be especially relevant to those government regulators and industry scientists who are striving to assess a backlog of inadequately tested chemicals, such as through the REACH program, as well as to scientists in government, industry, and academia who are seeking to implement pathway-based testing along the lines proposed in the 2007 NRC report cited above. REACH entails, among other components, a thorough assessment of toxicological dossiers on individual chemicals, which is best accomplished via an evidence-based approach. Pathway-based testing is developing too rapidly to be amenable to current validation approaches; evidence-based approaches can fill this void, especially where the reference standard shifts from high dose animal studies to human biology.



The benefits of translating evidence-based approaches from medicine to toxicology have been recognized since 2005. Experience since then has shown that this translation will not reach a critical mass in practice unless interested stakeholders join together in a focused, well-funded effort. The January 2012 workshop was an important step in moving forward.

References

- Abhyankar, L. N., Jones, M. R., Guallar, E., et al. (2011). Arsenic exposure and hypertension: a systematic review. *Environ Health Perspect* 120, 494-500.
- Adami, H.-O., Berry, C. L., Breckenridge, C. B., et al. (2011). Toxicology and epidemiology: improving the science with a framework for combining toxicological and epidemiological evidence to establish causal inference. *Toxicol Sci* 122, 223-234.
- Bhattacharya, S., Zhang, Q., Carmichael, P., et al. (2011). Toxicity testing in the 21st century: defining new risk assessment approaches based on perturbations of intracellular toxicity pathways. *PLoS ONE* 6, e20887.
- Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69, 89-95.
- Conrad, J. W. and Becker, R. A. (2010). Enhancing credibility of chemical safety studies: an emerging consensus on key assessment criteria. *Environ Health Perspect* 119, 757-764.
- Covey, S. R. (1989). *The Seven Habits of Highly Effective People: Restoring The Character Ethic*. New York: Simon and Schuster.
- Dieterle, F., Perentes, E., Cordier, A., et al. (2010a). Urinary clusterin, cystatin C, beta2-microglobulin and total protein as markers to detect drug-induced kidney injury. *Nat Biotechnol* 28, 463-469.
- Dieterle, F., Sistare, F., Goodsaid, F., et al. (2010b). Renal biomarker qualification submission: a dialog between the FDA-EMA and Predictive Safety Testing Consortium. *Nat Biotechnol* 28, 455-462.
- Griesinger, C., Hoffmann, S., Kinser, A., et al. (2009). Proceedings of the 1st international forum towards evidence-based toxicology. *Preface. Hum Exp Toxicol* 28, 83-86.
- Guyatt, G., Cairns, J., Churchill, D., et al. (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 268, 2420-2425.
- Guzelian, P. S., Victoroff, M. S., Halmes, N. C., et al. (2005). Evidence-based toxicology: a comprehensive framework for causation. *Hum Exp Toxicol* 24, 161-201.
- Hamburg, M. (2011). Advancing regulatory science. *Science* 331, 987.
- Harpur, E., Ennulat, D., Hoffman, D., et al. (2011). Biological qualification of biomarkers of chemical-induced renal toxicity in two strains of male rat. *Toxicol Sci* 122, 235-252.
- Hartung, T. (2009). Food for thought ... on evidence-based toxicology. *ALTEX* 26, 75-82.
- Hartung, T. (2010). Evidence-based toxicology – the toolbox of validation for the 21st century? *ALTEX* 27, 253-263.
- Hoffmann, S. and Hartung, T. (2005). Diagnosis: toxic! – trying to apply approaches of clinical diagnostics and prevalence in toxicology considerations. *Toxicol Sci* 85, 422-428.
- Hoffmann, S. and Hartung, T. (2006). Toward an evidence-based toxicology. *Hum Exp Toxicol* 25, 497-513.
- Hoffmann, D., Adler, M., Vaidya, V. S., et al. (2010). Performance of novel kidney biomarkers in preclinical toxicity studies. *Toxicol Sci* 116, 8-22.
- Hoffmann, S. (2012). Kick-off of the Evidence-based Toxicology Collaboration Europe. *ALTEX* 29, 456.
- Hoffmann, S., Stephens, M. L., and Hartung, T. (in press). Evidence-based toxicology. *Encyclopedia of Toxicology* (3rd edition). Elsevier.
- Jennings, C. A. (2006). *Roberts Rules for Dummies*. Hoboken: Wiley Publishing Company.
- Judson, R., Kavlock, R., Martin, M., et al. (2013). Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *ALTEX* 30, 51-66.
- Keller, D. A., Juberg, D. R., Catlin, N., et al. (2012). Identification and characterization of adverse effects in 21st century toxicology. *Toxicol Sci* 126, 291-297.
- Kotter, J. P. (1996). *Leading Change*. Boston, MA: Harvard Business School Press.
- Mauil, E. A., Ahsan, H., Edwards, J., et al. (2012). Evaluation of the association between arsenic and diabetes: a National Toxicology Program workshop review. *Environ Health Perspect* 120, 1658-1670. <http://dx.doi.org/10.1289/ehp.1104579>
- Mayer, D. (2004). *Essential Evidence-Based Medicine*. New York: Cambridge University Press.
- Navas-Acien, A., Silbergeld, E. K., Streeter, R. A., et al. (2006). Arsenic exposure and type 2 diabetes: A systematic review of the experimental and epidemiological evidence. *Environ Health Perspect* 114, 641-648.
- Navas-Acien, A., Guallar, E., Silbergeld, E. K., et al. (2007). Lead exposure and cardiovascular disease: A systematic review. *Environ Health Perspect* 115, 472-482.
- National Research Council – NRC (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington D.C.: National Academies Press.
- National Research Council (2011). *Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde*. Washington, DC: National Academies Press. <http://www.nap.edu/catalog/13142.html>
- Ozer, J. S., Dieterle, F., Troth, S., et al. (2010). A panel of urinary biomarkers to monitor reversibility of renal injury and a serum marker with improved potential to assess renal function. *Nat Biotechnol* 28, 486-494.
- Schreider, J., Barrow, C., Birchfield, N., et al. (2010). Enhancing the credibility of decisions based on scientific conclusions: transparency is imperative. *Toxicol Sci* 116, 5-7.
- Seed, J., Carney, E., Corley, R., et al. (2005). Overview: using mode of action and life stage information to evaluate the human relevance of animal toxicity data. *Crit Rev Toxicol* 35, 663-672.
- Silbergeld, E. and Scherer, R. (2013). Evidence-based Toxicology: strait is the gate, but the road is worth taking. *ALTEX* 30, 67-73.



- Starr, M. I. (2009). The origins, evolution and future of The Cochrance Database of Systematic Reviews. *Int J Technol Assess Health Care* 25, Suppl 1, 182-195.
- Sutton, P., Woodruff, T. J., Vogel, S., et al. (2011). Conrad and Becker's "10 Criteria" fall short of addressing conflicts of interest in chemical safety studies. *Environ Health Perspect* 119, a506-a507.
- US Food and Drug Administration (2010). Guidance for industry: qualification process for drug development tools. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf>
- Vaidya, V. S., Ozer, J. S., Dieterle, F., et al. (2010). Kidney injury molecule-1 outperforms traditional biomarkers of kidney injury in preclinical biomarker qualification studies. *Nat Biotechnol* 28, 478-485.
- Welytok, J. G. (2007). *Nonprofit Law and Governance for Dummies*. Hoboken, NJ: Wiley.
- Woodcock, J., Buckman, S., Goodsaid, F., et al. (2011). Qualifying biomarkers for use in drug development: a U.S. Food and Drug Administration overview. *Expert Opin Med Diagn* 5, 369-374.
- Woodruff, T. J., Sutton, P., and The Navigation Guide Work Group (2011). An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. *Health Affairs* 30, 931-937.
- Yu, Y., Jin, H., Holder, D., et al. (2010). Urinary biomarkers trefoil factor 3 and albumin enable early detection of kidney tubular injury. *Nat Biotechnol* 28, 470-477.
- Zurlo, J. (2011). Evidence-based Toxicology Collaboration kick-off meeting. *ALTEX* 28, 151.

Acknowledgements

The Evidence-based Toxicology Collaboration would like to thank the following in connection with its January 2012 workshop: Robert Kavlock and the Environmental Protection Agency for hosting the workshop; Marilyn Principe of the Johns Hopkins Center for Alternatives to Animal Testing for logistical support; Kellyn Betts for summarizing the comments from the workshop; the American Chemistry Council, American Cleaning Institute, ExxonMobil, and the transatlantic think tank for toxicology (t⁴) sponsored by the Doerenkamp-Zbinden Foundation for financial support; and the Society of Toxicology for its endorsement of the workshop.

Correspondence to

Martin Stephens, PhD
Center for Alternatives to Animal Testing
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street, W7032
Baltimore, MD, 21205
USA
e-mail: mstephen@jhsph.edu