# UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
## WASHINGTON D.C., 20460

OFFICE OF
CHEMICAL SAFETY AND
POLLUTION PREVENTION

October 28, 2014

## MEMORANDUM

**SUBJECT:** Transmittal of Meeting Minutes of the FIFRA Scientific Advisory Panel Meeting held July 29-30, 2014 on" New High Throughput Methods to Estimate Chemical Exposure"

**TO:** David Dix, Ph.D., Director
Office of Science Coordination and Policy

**FROM:** Fred Jenkins, Jr., Ph.D., Designated Federal Official
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

**THRU:** Laura Bailey, M.S., Executive Secretary
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

Attached, please find the meeting minutes of the FIFRA Scientific Advisory Panel open meeting held in Arlington, VA on July 29-30, 2014. This report addresses a set of scientific issues associated with "New High Throughput Methods to Estimate Chemical Exposure."

Enclosure

cc:
Jim Jones
Louise Wise
Jack Housenger
William Jordan
Margie Fehrenbach
Yu-Ting Guilaran
Robert McNally
Donald Brady
Jacqueline Mosby
Dana Vogel
Susan Lewis
Richard Keigwin
David Dix
Laura Bailey
Tina Bahadori
Rusty Thomas
Jim Cowles
Craig Barber
Alan Dixon
Peter P. Egeghy
Kristin Isaacs
Steven Knott
Woodrow Setzer
John Wambaugh
Cathy Milbourn
Linda Strauss
OPP Docket

FIFRA Scientific Advisory Panel Members
James McManaman, Ph.D.
Dana Boyd Barr, Ph.D.
Kenneth Delclos, Ph.D.
Marion Ehrich, Ph.D., D.A.B.T., A.T.S.
David Jett, Ph.D.

FQPA Science Review Board Members
James Chen, Ph.D.
Mark Cronin, Ph.D.
Panagiotis Georgopoulos, Ph.D.
William Hayton, Ph.D.
Peter Macdonald, D.Phil., P.Stat.
Cheryl Anne Murphy, Ph.D.
Thomas Potter, Ph.D.
Daniel Schlenk, Ph.D.

# FIFRA Scientific Advisory Panel Minutes No. 2014-03

# A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding New High Throughput Methods to Estimate Chemical Exposure

## July 29-30, 2014
## FIFRA Scientific Advisory Panel Meeting
## Held at the
## EPA Conference Center
## Arlington, VA

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) is a Federal advisory committee operating in accordance with the Federal Advisory Committee Act and established under the provisions of FIFRA as amended by the Food Quality Protection Act (FQPA) of 1996. The FIFRA SAP provides scientific advice, information, and recommendations to the EPA Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The meeting minutes represent the views and recommendations of the FIFRA SAP and do not necessarily represent the views and policies of the EPA or of other agencies in the Executive Branch of the Federal government. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use. The meeting minutes do not create or confer legal rights or impose any legally binding requirements on the EPA or any party.

# TABLE OF CONTENTS

## 1. NOTICE

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) is a Federal advisory committee operating in accordance with the Federal Advisory Committee Act and established under the provisions of FIFRA as amended by the Food Quality Protection Act (FQPA) of 1996. The FIFRA SAP provides advice, information, and recommendations to the Agency Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The Panel serves as the primary scientific peer review mechanism of the Environmental Protection Agency (EPA), Office of Pesticide Programs (OPP), and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. FQPA Science Review Board members serve the FIFRA SAP on an *ad hoc* basis to assist in reviews conducted by the FIFRA SAP. The meeting minutes have been written as part of the activities of the FIFRA SAP.

In preparing the meeting minutes, the FIFRA SAP carefully considered all information provided and presented by EPA, as well as information presented in public comment. The minutes represent the views and recommendations of the FIFRA SAP and do not necessarily represent the views and policies of the EPA, nor of other agencies in the Executive Branch of the Federal government. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use. The meeting minutes do not create or confer legal rights or impose any legally binding requirements on EPA or any party.

The meeting minutes of the July 29-30, 2014 FIFRA SAP meeting held to consider and review scientific issues associated with "New High Throughput Methods to Estimate Chemical Exposure" were certified by James McManaman, Ph.D., FIFRA SAP Session Chair, and Fred Jenkins, Ph.D., FIFRA SAP Designated Federal Official, on October 28, 2014. The minutes were reviewed by Laura E. Bailey, M.S., FIFRA SAP Executive Secretary. The minutes are publicly available on the SAP website (http://www.epa.gov/scipoly/sap/) under the heading of "Meetings" and in the public e-docket, Docket No. EPA-HQ-OPP-2014-0331, accessible through the docket portal: http://www.regulations.gov.

Further information about FIFRA SAP reports and activities can be obtained from its website at http://www.epa.gov/scipoly/sap/. Interested persons are invited to contact Fred Jenkins, Ph.D., SAP Designated Federal Official, via e-mail at jenkins.fred@epa.gov.

# SAP Minutes No. 2014-03

# A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

# New High Throughput Methods to Estimate Chemical Exposure

# July 29-30, 2014
## FIFRA Scientific Advisory Panel Meeting
## Held at
## One Potomac Yard
## Arlington, Virginia

*James L. McManaman*

**James L. McManaman, Ph.D.**
**FIFRA SAP Session Chair**
**FIFRA Scientific Advisory Panel**

*Fred Jenkins*

**Fred Jenkins, Jr., Ph.D.**
**Designated Federal Official**
**FIFRA Scientific Advisory Panel Staff**

**Date:** October 28, 2014

**Date:** October 28, 2014

## 2. PANEL ROSTER

**FIFRA SAP Chair**

**James L. McManaman, Ph.D.**
Professor and Chief
Section of Basic Reproductive Sciences
Department of Obstetrics and Gynecology and Physiology
and Biophysics
University of Colorado Denver
Aurora, CO

**Designated Federal Official**

**Fred Jenkins, Jr., Ph.D.**
U.S. Environmental Protection Agency
Office of Science Coordination and Policy
FIFRA Scientific Advisory Panel
EPA East Building, MC 7201M
1200 Pennsylvania Avenue, NW
Washington, DC 20460
Phone (202) 564-3327
Fax (202) 564-8382
jenkins.fred@epa.gov

**FIFRA Scientific Advisory Panel Members**

**Dana Boyd Barr, Ph.D.**
Research Professor
Department of Environmental and Occupational
Health
Rollins School of Public Health
Emory University
Atlanta, GA

**Kenneth Delclos**, **Ph.D.**
Research Pharmacologist
Division of Biochemistry Toxicology
National Center for Toxicological Research
U.S. Food and Drug Administration
Jefferson, AR

**Marion Ehrich, Ph.D., D.A.B.T., A.T.S.**
Professor of Pharmacology and Toxicology
Department of Biomedical Science and
Pathobiology
Virginia-Maryland College of Veterinary
Medicine
Virginia Polytechnic Institute and State
University
Blacksburg, VA

**David Jett, Ph.D.**
Program Director
National Institute of Neurological Disorders and
Stroke
National Institutes of Health
Bethesda, MD

### 3. List of Commonly Used Acronyms and Abbreviations

| | |
|---|---|
| ACToR: | EPA's Aggregated Computational Toxicology Repository |
| AIC: | Akaike Information Criterion |
| | |
| BW: | Body weight |
| CDC: | the United States Centers for Disease Control |
| CDR: | EPA Chemical Data Reporting (formerly IUR) |
| CL | Clearance |
| CLint | Intrinsic Hepatic Clearance |
| CLr | Renal Clearance |
| CPCat: | Chemical-Product Categorization database |
| CPCPdb: | Consumer Product Chemical Profiles database |
| Css: | plasma Concentration after Steady-State is reached due to repeated exposures |
| DSSTox: | Distributed Structure-Searchable Toxicity Database |
| EDSP: | the Endocrine Disrupter Screening Program |
| EDSTAC: | Endocrine Disruptors Screening and Testing Advisory Committee |
| EEC: | Expected Environmental Concentration |
| EPA: | the United States Environmental Protection Agency |
| EPI Suite: | EPA's Estimation Program Interface Suite |
| ExpoCast: | EPA's Exposure ForeCast prioritization research program |
| EXAMS: | EPA's Exposure Analysis Modeling System |
| FDA: | the United States Food and Drug Administration |
| FIFRA: | Federal Insecticide, Fungicide, and Rodenticide Act |
| fup: | fraction of chemical that is unbound in plasma when in the presence of endogenous levels of plasma protein |
| GFR | Glomerular Filtration Rate |
| HT: | High-throughput |
| HTE: | High-throughout Exposure |
| HTTK | High Throughput Toxicokinetics () |
| HTPBTK | Physiologically Based High Throughput Toxicokinetics |
| HTS: | High-throughput screening |
| IVIVE: | *in vitro-in vivo* extrapolation |
| KABAM: | EPA's Kow (based) Aquatic BioAccumulation Model |
| KOW: | Octanol-Water partition coefficient (also called lipophilicity or Hydrophobic |
| KNIME | Konstanz Information Miner |
| NHANES | National Health and Nutrition Examination Survey |
| RTK | Reverse Toxicokinetics |
| SEEM | Systematic Empirical Evaluation of Models |
| SHEDS | Stochastic Human Exposure and Dose Simulation |

## 4. INTRODUCTION

On July 29-July 30, 2014 the US EPA Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel (FIFRA SAP) met in Crystal City, VA to consider and review scientific issues associated with "New High Throughput Methods to Estimate Chemical Exposure". EPA is interested in using a methodology called "ExpoCast" to predict human health and wildlife exposures to chemicals. EPA specifically proposed using this technology to screen for pesticides and other chemicals that may potentially cause endocrine disruption. The Panel provided recommendations to EPA regarding this proposal. Specifically, the Panel addressed the EPA's charge to them on questions concerning the following 3 topic areas: 1) The Systematic Empirical Evaluation of Models (SEEM) Framework for Exposure, 2) High Throughput Toxicokinetics (HTTK) and Reverse Toxicokinetics (RTK), and 3) Future Direction. Opening remarks at the meeting were provided by Dr. David Dix, Director, Office of Science Coordination and Policy.

US EPA presentations were provided by the following EPA staff (listed in alphabetical order):

Craig Barber
Alan Dixon
Peter P. Egeghy
Kristin Isaacs
Steven Knott
Woodrow Setzer
John Wambaugh

## 5. PUBLIC COMMENTERS

*Oral public comments are listed in the order presented during the meeting*:

Christopher J. Borgert of Applied Pharmacology and Toxicology, Inc. on behalf of the Endocrine Policy Forum
Paul S. Price of the Dow Chemical Company on behalf of the Endocrine Policy Forum
Michael Bartels of the Dow Chemical Company on behalf of the Endocrine Policy Forum
Richard A. Becker of American Chemistry Council on behalf of the Endocrine Policy Forum

*Written statements were provided by (listed in alphabetical order):*

Patricia L. Bishop and Kristie Sullivan respectively on behalf of the People for the Ethical Treatment of Animals and the Physicians Committee for Responsible Medicine
Crystal Lake, a private citizen, on behalf of herself
Ellen Mihaich on behalf of the Endocrine Policy Forum
Sacoby Wilson of the University of Maryland on behalf of himself

## 6. **OVERALL SUMMARY**

The Panel was charged with advising EPA on the topic areas of: 1) The Systematic Empirical Evaluation of Models (SEEM) Framework for Exposure, 2) High Throughput Toxicokinetics (HTTK) and Reverse Toxicokinetics (RTK), and 3) Future Direction.

In regard to the SEEM Framework for Exposure the Panel commended, the Agency for making significant progress in advancing the SEEM framework to assess risks of chemical exposures. The Panel recommended that the SEEM framework be used within the context of a screening program which evaluates prediction via conservative/worst scenarios (i.e. assessing exposures in the most sensitive subpopulations including children). Additionally the Panel recommended that before the SEEM framework is implemented into the EPA's Endocrine Disruptor Screening Program (EDSP), it should undergo further modifications to reduce the uncertainty associated with it. The Panel also advised the Agency to collaborate with the Center for Disease Control in an effort to determine measurements that would be beneficial.

The Panel concurred with the Agency's overall approach to evaluating HTTK and RTK modeling. However, they explained several challenges and limitations associated with this approach. For example, there are insufficient data, and the testing fails to represent the diversity of chemicals. The Panel also recommended that the proposed models be expanded beyond evaluating oral exposure to include pulmonary and dermal exposure routes. Additionally the Panel recommended further enhancement of the HTTK and RTK models that entails more detailed documentation and description of their underpinnings.

Regarding future direction the Panel advised the Agency to adopt the term "bioactivity" as a means of defining biological activity as opposed to using the term "activity". They also recommended that the Agency expand its dimensions for evaluating exposure to include measurements of pervasiveness, persistence, severity, and efficacy. Concerning ecological exposure the Panel agreed that the Agency has taken a good initial step by coupling the SHEDS and EXAMS models to assess exposure to ecosystems. However, the Panel expressed caution that utilizing aggregated water sampling for environmental exposure may potentially underestimate risks in vulnerable areas. The Panel noted that further attention should be directed to residential and non-residential exposures that are influenced via human behavior/activities and attributes of microenvironments in which the human activities occur. Lastly, as a recommendation for future direction, the Panel encouraged the Agency to more in depthly consider the application of additional QSAR methodologies to support their efforts.

## 7. EXECUTIVE SUMMARY OF PANEL DISCUSSION AND RECOMMENDATIONS

**Charge Question 1.1**. In the absence of sufficient exposure information to estimate exposure for the majority of the chemicals of interest in the Endocrine Disruptor Screening Program (EDSP) using EPA's historical low-throughput methods, please comment on whether the Systematic Empirical Evaluation of Models (SEEM) approach is scientifically sound and suitable for using High-throughput Exposure (HTE) methods to estimate relative levels of chemical exposures and the associated uncertainty of these estimates for consideration in a prioritization approach.

*Panel Summary*

Panel members acknowledged that EPA has made substantial progress in developing HTE methods and that the SEEM approach appears to be scientifically sound and suitable for HTE methods to assess relative risk of chemical exposures from diverse groups of chemicals. It was emphasized that the primary areas of progress were in problem formulation and conceptual model development. These are critical first steps and are consistent with tiered modeling approaches that are the *de facto* standard for regulatory risk assessment. Development of a model evaluation framework for SEEM and efforts to use forward and reverse modeling to compare measured and modeled data were also notable accomplishments.

There was also general agreement that further effort in measuring and minimizing uncertainty within the SEEM framework is needed prior to implementation in the EDSP or other programs. To achieve this result, it was noted that the White Paper correctly stated that further large scale efforts to obtain data for model evaluation are needed. The data presently available appear insufficient to evaluate model performance systematically.

The Panel expressed concern that the White Paper in many cases did not provide sufficient detail to thoroughly evaluate the SEEM methodology. A revision of the White Paper was recommended to include explanations and data presented at the meeting by EPA staff. There was consensus that presentations provided during the public SAP meeting were generally of high quality and clarified important points in the SEEM analysis framework.

It was recommended that SEEM, as a screening program, represents quality science adequate for testing or validating predictions using conservative "worst-case" scenarios. For example, splitting categories that focus on sensitive groups such as children or splitting pesticide use for urban/indoor and agricultural use would be very beneficial. Using worst-case scenarios as empirical tests of model validity is consistent with a "screening" procedure. As a screening process, the science should err more toward Type I error (false positive). It was also recommended that HTE may not only be used simultaneously with High Through Put (HTP) bioassays for screening, but may provide environmentally relevant dosages for Tier I studies (or additional HTP bioassays) that would avoid overt toxicity that confounds endocrine responses *in vivo*.

**Charge Question 1.2**. Please suggest the most important steps for EPA that could decrease uncertainties and increase confidence in using the HTE approach to predict exposures in various demographic groups (e.g., young children, women of child bearing age) for large numbers of chemicals.

*Panel Summary*

The Panel identified several limitations that need addressing regarding the use of NHANES datasets for calibration and validation. In many instances, parameters were not consistently measured across NHANES cycles. Thus, true exposures may have been underestimated because of the selected species measured and a lack of consideration for pharmacokinetic information of derived NHANES-equivalents. Other considerations included the lack of vulnerable groups evaluated in NHANES, and the lack of information on the temporality and variability within measurements. The Panel recommended that EPA collaborate with CDC to implement measurements useful for evaluation of the SEEM approach. The Panel further recommended that the Agency consider which parameters differ among various demographic groups and which need special attention. For example, while the pharmacokinetics may be similar among groups, the dose (active ingredient/body weight) may differ. Specifically, these areas should be targeted to determine if more parameters should be considered for sensitive groups. Lastly, providing examples of best-case and worst-case scenarios of validation and calibration in the White paper is warranted.

**Charge Question 2.1** In the absence of sufficient empirical toxicokinetic information for the thousands of EDSP relevant chemicals, please comment on the approach of using HTTK, HTPBTK, IVIVE, and RTK for estimating chemical TK to provide an administered dose context to the concentrations showing bioactivity in the endocrine-related HTS assays.

*Panel Summary*

The Panel was impressed by the approach taken by the EPA scientists to tackle this challenging issue. Overall, the Panel agreed that the Agency approach was in the right direction, and that there appears to be a large body of literature available that supports the approach r. The main assumption, that the free concentration in water at the biological target will be bioactive, is valid. However, there are limitations and constraints that must be acknowledged, which have been outlined by the Agency, the Endocrine Policy Forum and the Panel. The framework setup by these modeling efforts is excellent and will allow for further model refinement and validation once more chemical datasets (other than pharmaceuticals) become available.

**Charge Question 2.2** A comparison of the HTTK-predicted steady-state blood concentrations with *in vivo* values from the literature suggests that the overall correlation is low, but that the discrepancy between the two can be predicted using a combination of chemical properties, quantitative structure activity relationships (QSAR), and cutoffs from the HTTK assays. Please comment on: a) how well this approach characterizes the uncertainty in the steady-state blood concentrations, and b) whether the identification of chemical classes that need additional TK investigation is useful in a chemical prioritization or initial screening context.

---

*Panel Summary*

The Panel noted that the Agency has made progress in identifying uncertainty in the steady state blood concentrations (Css). The Panel detailed several limitations to the Agency's approach that if addressed would allow the uncertainty to be better characterized. The limitations included: 1) a deficiency in defining the underlying models, 2) a lack of identifying the domains of models and *in vitro* systems. (consequently making it difficult to determine if poor prediction could be classed as "reliable"), 3) the quality of the data modeled, and 4) the potential uncertainties associated with factors such as extrahepatic metabolism, enterohepatic recirculation, unusual binding affinity for one tissue, transporters, and poor absorption (none of these factors have not been fully evaluated).

**Charge Question 2.3** Please comment on whether the assumptions made in these models are appropriate given the current state of the science and data limitations.

*Panel Summary*

The models proposed were standard compartmental (HTTK) and physiologically based (HTPBTK) toxicokinetic models, appropriate for the intended use. The implicit assumptions of the models were that the mammalian body could be represented as one or more "compartments", and that chemical movement among compartments could be described by linear differential equations. Although the chemical input was by oral route, the models could be adapted to other inputs such as pulmonary and dermal, and could be adapted to include a separate compartment for the site of toxicity (i.e. target organ). A compartment may be characterized by its capacity to accumulate or hold a chemical (Volume of Distribution, V), and the assumption of linearity (i.e., independent of the amount of chemical in the compartment) was appropriate. A further assumption associated with compartmentalization was that after entrance to a compartment, the chemical distribution throughout the compartment was instantaneous (i.e., that there were no concentration gradients within a compartment). This assumption is inherent in the concept of a compartment, and was considered appropriately. Finally, it was assumed that the V value for a chemical could be calculated using a published (ref. 89 of the Agency's White Paper) tissue model comprised of water, neutral lipids, neutral and acidic phospholipids, and proteins combined with the use of chemical properties, such as lipophilicity, binding to phospholipid membranes, pKa and the unbound fraction in blood plasma. Ref. 89 provided compelling results, which validated and supported the appropriateness of its use.

In addition to compartments, the models were comprised of model parameters (rate constants, k, and clearance constants, CL) whose numerical values controlled the rates of chemical movement into and out of the compartments. For the oral route of administration, all the administered dose was assumed to be absorbed from the gastrointestinal (GI) tract, an assumption not appropriate for many chemicals. The rate constant for absorption was assumed to have a value of $1 \text{ h}^{-1}$, which was considered to be appropriate. The rate of distribution of chemical among compartments was assumed to be controlled by the blood flow to the tissues associated with the compartment. This assumption was considered generally appropriate although it may not be applicable to chemicals that do not readily permeate the membranes of cells because of their specific chemical characteristics (i.e. high molecular weight or high polarity).

To enable high throughput screening of a large number of chemicals, it was assumed that the free fraction (fup) of chemical in the blood plasma/serum could be quantified *in vitro* using rapid equilibrium dialysis and that the intrinsic hepatic clearance (CLint) could also be quantified *in vitro*. These approaches were appropriate approaches, and have been validated by others. Complications were noted,

however, when fup was less than 1% and when hepatic metabolism activity was low. When fup was less than 1%, random, draws were selected from a uniform distribution ranging from 0% to 1%, which appeared to be appropriate and superior to simply using a value of 0.5%. The Agency assumed that the CLint value for a chemical was concentration independent. This assumption appeared to be appropriate as long as the concentration of chemical in the body was low as which might be expected from an environmental exposure. It was also assumed that metabolites of a chemical could be ignored although they could be important contributors to toxicity and modulators of parent chemical toxicokinetics. The renal clearance (CLr) was assumed to be the product of fup and the glomerular filtration rate, GFR: CLr = fup x GFR. With this assumption, renal tubular reabsorption and secretion were ignored, which would not be appropriate for many chemicals.

A log-normal distribution was generally used for the Monte Carlo (MC) simulations; while this was considered appropriate, exploration of other distributions was recommended to increase confidence in the MC simulations. A final suggestion by the Panel was to perform a parameter sensitivity analysis to assess whether it was important to have highly accurate values for a particular parameter in order to obtain acceptable prediction of internal dose metrics.

**Charge Question 2.4** Please suggest the most important steps EPA should take to improve the various kinetic models to provide rapid and cost-effective predictions for large number of chemicals.

*Panel Summary*

The Panel suggested that the Agency find the cause(s) for the poor capacity of the kinetic models to predict steady-state plasma concentrations from the literature, Fig. 28. The Panel was concerned that the lack of correlation displayed in Fig. 28 of the White Paper stemmed from model deficiencies that might be correctable. Several suggestions were made. One deficiency was the assumption that renal clearance (CLr) was the product of the free fraction of chemical in plasma (fup) multiplied by the GFR; incorporation of active renal tubular secretion and tubular reabsorption may provide feasible improvement to high-throughput prediction of CLr values. Another suggestion was to incorporate a model to predict oral bioavailability, rather than assume that it was 100% for all chemicals. The Panel further suggested that refinements be sought in the methods for more accurate determination of fup when fup was less than 1%, and for more accurate determination of intrinsic hepatic clearance for low-hepatic-clearance chemicals. Also suggested was the addition of dermal and pulmonary routes of entry of chemicals to the HTTK and the HTPBTK models. To increase confidence in the prediction capabilities of the TK models, use of additional animal data for model validation was suggested. As the TK models incorporate information from *in vitro* assays and other *in silico* models, KNIME technology was suggested as a useful tool for combining the output from multiple sources. KNIME is the Konstanz Information Miner, which is an open source data analytics, reporting and integration platform. KNIME integrates various components for machine learning and data mining through its modular data pipelining concept (KNIME, 2014).

To assist users of the TK models and the models that provide input to them (e.g. the tissue partitioning model), the Panel suggested that it would be helpful to have a more extensive description and documentation of the models than was provided in the White Paper. Equations and algorithms should be explicitly stated, along with performance statistics and applicability domains. Model descriptions could be based around OECD Principles for the Validation of QSARs. Publication of the models in peer-reviewed journals was encouraged and release of the algorithms in R computer programming language was suggested.

**Charge Question 2.5** Please comment on the approaches presented in the present White Paper for comparing RTK-derived endocrine-active dose ranges with exposure predictions from ExpoCast. Discuss the strengths and limitations of these comparisons, and whether this or other approaches are suitable for clearly distinguishing chemicals with higher predicted doses from chemicals with lower predicted RTK adjusted doses.

*Panel Summary*

When compounds of concern are plotted against the "raw" ToxCast AC50 values there is no pattern. However, when these compounds are plotted against RTK (Reverse ToxicoKinetic) estimated oral equivalents, a pattern emerged that distinguished compounds with higher RTK oral equivalents from those with lower oral equivalents. These RTK oral equivalents were developed using a three-compartment model that estimates steady state concentrations in plasma using plasma binding and hepatocyte clearance assays and assumes linear relationships and steady states for daily dose and oral equivalent in *in vitro/in vivo* extrapolations. When these assumptions are acceptable (for the compounds and concentration ranges considered) and uncertainties are taken into consideration, this appears to be a method that can distinguish the higher predicted dose range from the lower predicted dose range, although middle-range doses may not be clearly distinguished from each other.

More generally, given the approximate nature of HT screening/ranking, the evolving approaches presented in the White Paper for comparing RTK-derived endocrine-active dose ranges with exposure predictions from ExpoCast should be considered generally appropriate, taking into account that uncertainties in both "forward" and "reverse" (RTK-derived) intake estimates span multiple orders of magnitude. It should be stated explicitly that comparisons must be limited to situations where exposures are low and follow regular repetitive patterns that would support the assumptions (linearity, steady state) incorporated in the HT models.

**Charge Question 3.1** Please comment on the three key areas including whether there are other areas that are of equal or higher priority to support an integrated activity/hazard and exposure based prioritization and screening approach within the EDSP or other chemical programs.

*Panel Summary*

The Panel advised that EPA use the term "bioactivity" instead of activity in defining the biological activity of a chemical to keep the terms discreet. Several dimensions of exposure should be described with the HTE analysis including pervasiveness, persistence, severity and efficacy. The planned expanded use of a probabilistic HT exposure model that employs a basic "mechanistic foundation," i.e. SHEDS-HT in a framework that explicitly discriminates variabilities and uncertainties (e.g. via 2D-dimensional Monte Carlo analyses) will likely be more useful than the current 1-D approach.

**Charge Question 3.2** We propose applying the ExpoCast framework to ecological exposures using aggregated water monitoring data to evaluate the predictions of environmental fate and transport models. Please identify other data, models or environmental media that may be of greater value in the initial model calibration and uncertainty analysis.

*Panel Summary*

The Panel recommended that the proposed models were a positive movement by the Agency to address ecological exposure. In particular, the coupling of SHEDs to EXAMS was considered to be a good first step. However, panel members cautioned that using aggregated water sampling for environmental exposure as an "average" potential for exposure may underestimate exposures in areas prone to greater discharge and higher concentrations. While probabilistic estimates may be obtained using this method, a much simpler and more conservative approach would be to utilize "worst-case" scenarios as initial screens. This process has been recommended for screening "emerging compounds" for the State of California (Drews et al., 2012; Maruya et al., 2013). For a mechanism to assess monitoring data for human health for drinking water derived from recycled water, it was recommended that EPA consult the findings of a Science Advisory Panel convened by the State of California for identification of emerging contaminants for monitoring (Anderson et al. 2010; Anderson et al. 2012; Drewes et al.,2013; Maruya et al. 2013).

The Panel emphasized that alternatives or companions to monitoring data are concentration estimates derived by simulation modeling. A core EPA approach for pesticide simulations is use of the field scale model PRZM coupled to the EXAMS model. Thus, the Panel indicated that it may be useful to couple PRZM-EXAMS to SHED-PBTK. Other models that were recommended for use at larger spatial scales were SWAT (Soil Water Assessment Tool) and APEX (Agricultural Policy Extender Model). They were recently used in the USDA Conservation Effects Assessment Program National Assessment for Cropland (CEAP, 2014). Other suggestions were for the Agency to consider construction of simple agent based particle models (with chemical properties assigned to the "agents") to explore the various contaminant transmission pathways using environmental pathways in a GIS framework. These can then be used in pattern-oriented modeling (POM) to determine likely routes of transmission that can be evaluated against the aggregated water monitoring data.

**Charge Question 3.3** Please identify other relatively high-throughput sources of exposure-related information that should be included in these three key areas and suggest how this information would be used to help prioritize chemicals.

*Panel Summary*

Areas that requires further attention and ("distributional") characterization appropriate for HT modeling, are human behavior/activities and attributes of microenvironments in which the human activities take place (both residential and non-residential). These areas are critical in driving exposures as well as doses.

Building/improving distributions of human activities/behavior, appropriate for use with HT models, that account for variabilities associated with age, gender, urban and rural location, socioeconomic status, ethnic and cultural background, etc. is important for both "forward" exposure modeling and for proper interpretation of chemical biomonitoring data via "reverse" exposure modeling. Various other extant sources are available from federal and state agencies and from professional groups or associations.

Mining social media is an area that should be considered as a means for extracting patterns of human behavior, especially involving new products.

**Charge Question 3.4.** For the HTTK work going forward, please comment on additional studies that could be performed or approaches that could be taken to improve rapid and cost efficient predictions of TK for large numbers of chemicals.

*Panel Summary*

The Panel suggested that EPA consider and evaluate the applicability and usefulness of additional QSAR methods for developing estimates for absorption, distribution, metabolism, and excretion (ADME) parameters where currently default values/assumptions are used. The Agency should also consider this for cell cultures that aim: 1) to simulate biological processes and cellular interactions, and 2) to provide parameters in a manner more relevant to actual tissues and organs compared to single-cell cultures.

## DETAILED RESPONSES TO CHARGE QUESTIONS

**Question 1.1. In the absence of sufficient exposure information to estimate exposure for the majority of the chemicals of interest in the EDSP using EPA's historical low-throughput methods, please comment on whether the SEEM approach is scientifically sound and suitable for using high throughput exposure (HTE) methods to estimate relative levels of chemical exposures and the associated uncertainty of these estimates for consideration in a prioritization approach.**

The Panel noted that EPA has made substantial progress in developing HTE methods and that the SEEM approach appears to be scientifically sound and suitable for HTE methods to assess relative risks of chemical exposures from diverse groups of chemicals. It was emphasized that the primary areas of progress were in problem formulation and conceptual model development. These are critical first steps and are consistent with tiered modeling approaches that are the *de facto* standard for regulatory risk assessment. Development of a model evaluation framework for SEEM and efforts to use forward and reverse modeling to compare measured and modeled data were also notable accomplishments. These efforts are in agreement with the common methods for developing, assessing and evaluating most models.

There was also general agreement that further effort in measuring and minimizing uncertainty within the SEEM framework is needed prior to implementation in the EDSP or other programs. The essence of SEEM as described in the White Paper (pg. 13), is "use of rigorous statistical techniques to quantify uncertainties in HTE predictions using the limited data available". The key in this case is embedded in the last phrase, "limited data available". As with most models of this sort, quantifying and minimizing uncertainties requires measured data for an appropriately broad set of compounds. These data are used for calibration and validation and lead to parameter adjustment to minimize error and sensitivity analysis to determine which parameters have the greatest impact on model error. Error in this case is evaluated using a variety of well-established metrics (e.g. root-mean-square deviation (RMSE), percent bias (PBIAS)) to compare measured and simulated values. The magnitude of these metrics ultimately leads to decisions about model accuracy and ideally can be compared to consensus values that guide conclusions about model performance and suitability of use. This approach is by no means settled in science. Debate persists in nearly all fields regarding the magnitude of model performance parameters that constitute satisfactory performance. Moriasi et al. (2007) provided useful insight regarding model evaluation guidelines for watershed simulations. The models discussed were higher tier simulations than

SEEM. Nevertheless, the basic principles regarding the need for consensus criteria for "acceptable model performance" were relevant.

In summary, the science behind the framework appears sound but considerably more work is needed to assess and minimize uncertainty and determine probabilities of false negative or positive chemical classifications. The Panel concurs with the Agency's position that further large scale efforts to obtain data needed for model evaluation are needed in order to achieve this. Presently, available data appear insufficient to systematically evaluate model performance.

The Panel also expressed concern that the White Paper in many cases did not provide sufficient detail to thoroughly evaluate SEEM methodology. For example there was a lack of clarity in how the Akaike Information Criterion (AIC) criteria were used in model selection for both the SEEM-1 and SEEM-2 analysis. In SEEM-1 analysis, Table 2 (pg. 45) showed 22 AIC and various indices for model selection. The accompanying text (pg. 44) indicated that model 16 was selected for use. The Agency provided no explanation regarding why this model was selected over models 17, 18, and 20 which had smaller AIC values. In SEEM-2 analysis, the text indicated a five parameter model was selected based on the analysis shown in Figure 17 (pg. 55). Further, Figure 22 (pg. 62) showed an analysis of $R^2$ values for 5 models (i.e. those with 2-, 3-, 4-, 5- and 16-parameters). There were two issues identified by the Panel for which clarification was needed. First, the 16 parameter model appeared to have smaller $R^2$ values than the 4- and 5-parameter models. Thus, the 16-parameter model appeared to provide superior performance. Second, $R^2$ values among the 3-, 4- and 5-parameter models appeared to be close, and it can be concluded that these models fit the data equally well. Therefore, selection of the 5-parameter model did not appear to be consistent with the AIC approach (i.e. achieving the best model fit with the lowest number of parameters).

Other concerns expressed by the Panel focused on the White Paper's lack of clarity in descriptions of the use of Bayesian methods. It was noted that the explanation of Bayesian analysis in the White Paper (pg. 5) appeared to be more like classical hypothesis testing than Bayesian analysis. A better explanation of the benefits of Bayesian analysis was needed. For example, references to issues with the USGS/EPA water quality data (pg. 117) infers that Bayesian analysis will: 1) solve the unsolvable, 2) function as a viable substitute for incomplete and uncertain data, and 3) legitimately express real-world variability. Although Bayesian analysis gives results that are easier to communicate than classical frequentist results and Bayesian computational methods are now easier to implement with jags and similar software, with flat uninformative priors, the estimates and the measures of uncertainty will in the end be very close to frequentist results. An alternative (although not normal Bayesian practice) that was recommended involved utilizing informative priors to substitute for weak or uncertain data, with a justification for the choice of informative prior included in the discussion of model assumptions. Finally, one panel member observed that while the results of a Bayesian analysis are easier than p-values and confidence intervals to explain to non-statisticians, the definitions of the models and the prior distributions are not.

Of considerable interest to one panel member was the progression of utilizing the SEEM approach with other models such as SHEDs-HT which can also be linked with PBTK models. The linkage of these models will allow the utilization of non-human data such as that obtained using rodents for compounds already present in databases for FIFRA regulated pesticides as well as FDA approved pharmaceuticals and personal care products. Overall, combinations that integrate these models should eventually reduce uncertainty as refinements to the models occur and as additional data become available. As a screening program, it is recommended that the science used for SEEM test or validate predictions using conservative "worst-case" scenarios. For example, splitting categories that focus on sensitive groups such as children or splitting pesticide use for urban/indoor and agricultural use would be very beneficial.

Using worst-case scenarios as empirical tests of model validity is consistent with a "screening" procedure. As a screening process, the science should err more toward Type I error. It was recommended that HTE not only be used simultaneously with HTP bioassays for screening, but HTE may provide environmentally relevant dosages for Tier I studies (or additional HTP bioassays) that would avoid overt toxicity that confounds endocrine responses *in vivo*.

The Panel also focussed on implementation of the SEEM approach by forward and reverse prediction using physico-chemical property data and product use, release, and exposure characteristics. In addition, the Panel discussed application and evaluation of the Center for Disease Control's NHANES data for selected analytes used for model calibration. In exposure models physico-chemical property data were used to describe partitioning of chemicals within simulated environments and to also provide persistence estimates. Production, environmental release, and exposure data provided a population wide index of potential for steady-state exposure.

The White Paper stated that when available measured data for physico-chemical properties were used; however, for many parameters measured values were not readily available. In this case, QSAR models were used to predict properties using the Agency's EPI-SUITE program. The Panel agreed that the practice of using QSAR to fill data gaps is well established and, in part, is globally accepted. QSAR is particularly relevant to the prediction of physico-chemical properties, and KOWWIN (model which estimates the log octanol-water partition coefficient, log $K_{OW}$, of chemicals using an atom/fragment contribution method) within EPISuite is recognized as a good model and is freely available online. There was consensus that the embedded models in EPI-SUITE are scientifically sound and are suitable approaches for making prediction, however EPI-SUITE does not have a prediction algorithm for some key parameters like pKa. As suggested by the aforementioned factors, the panel members, it was recommended that EPA evaluate other models and consider which of the EDSP chemicals they are able to make useful predictions. The Panel cited Yu et al. (2010) which provides an overview of available QSAR models for predicting pKa of organic acids and nitrogen bases from molecular structure. This citation included an evaluation of the SPARC (SPARC Performs Automated Reasoning in Chemistry) QSAR model which is the model used by the Agency. Yu et al. (2010) found that the SPARC QSAR model did not perform as well as two other commercially available models.

Regarding use of QSARs in the White Paper, the Panel agreed that it would be good practice to fully describe these models and provide, if possible, information on applicability domains. One panel member observed that some effects and endpoints, e.g. tissue partitioning coefficients, have been problematic to model in the past, due to the paucity of high quality measurement data. Assuming good progress has been made in developing models for these endpoints, they would be very useful for other experts in related fields. The Panel also agreed that two questions that EPA may wish to consider include "how to further how to address the general concept of uncertainty in QSAR predictions?", and "how this adds to the overall uncertainty with the SEEM approach?"

Another area of concern was the use of QSAR in all cases to predict environmental degradation rates. Uncertainties associated with these estimates are unknown and may be high since QSAR models often perform poorly in predicting degradation rates (Raymond et al., 2001).

In the special case of pesticides, the Agency was encouraged to minimize use of QSAR for physioco-chemical property estimation and use available published measurements. A freely available source for many parameters is the FOOTPRINT Pesticide Properties Database (FOOTPRINT, 2014). FOOTPRINT data were compiled from European Union pesticide registration dossiers and generally considered to be of high quality. EPA also requires that registrants provide the Agency with measured environmental fate

property data. Thus, EPA scientists likely have access to high quality data for the numerous (hundreds) active ingredients in commercial use.

The Agency used many sources for data on chemical production, environmental release and exposure. In particular there was significant utilization of the EPA ACToR database system. Generally, the Panel advised that management of databases and updates to data require a more rigorous oversight since there does not seem to be a consistent plan or QA/QC criteria for data use in particular for environmental monitoring data from internet sources. For example, a better plan of how ACToR databases will be integrated with USGS/EPA NWQMC appears to be needed. Further if reverse kinetics are to be used from monitoring data, then sound science with regard to analytical chemistry and database use is strongly recommended. The Agency was also encouraged to explore alternate sources of pesticide use data such as recent compilations by the US Geological Survey. Using commercially available sales data and USDA-National Agricultural Statistics Service pesticide use estimates, Stone (2013) compiled use data for 459 compounds at the county level across the continental USA for the period 1992 through 2009. These county-level estimates provide a means to evaluate potential exposure "hotspots" and infer potential for exposure. Another national scale assessment by USDA-ERS evaluated use patterns of pesticides on 21 selected crops (Fernandez-Cornejo et al., 2014). Linking these estimates to where crops are produced appears to have the potential to improve precision of potential exposure estimates at least to pesticides used in crop production.

Finally, it was noted that SEEM analysis relied heavily on CDC-NHANES data, for model calibration and generation of uncertainty estimates. This data set represents the state-of-the-art in biomonitoring to assess human chemical exposure. However, given the complexity and cost of this type of monitoring, it is also not surprising that data are limited in scale and scope. The following were further points raised by the Panel concerning the NHANES data:

a. No samples were collected from children under the age of six; thus there are no exposure data for this highly important and vulnerable age group.

b. Across all samples that were collected and analyzed there was a high degree of left-censoring of the data, i.e. measurements were below the analytical limits of detection (LOD). For about 25% of targeted analytes there were no detections. Thus, it is unknown whether exposure occurred. In the White Paper in most cases it is unclear how LODs were handled in SEEM analyses. Some clarifications were provided by the Agency to the Panel during the public meeting. It was indicated that a lognormal distribution of values to the left of detection limits was assumed. One panel members noted that the lognormal model is inadequate given the large number of values below the LOD for many NHANES parameters. It was noted that a 0-lognormal mixture may be appropriate in some cases but impossible to fit with many LODs in data sets. A satisfactory solution would be to require better data with a more appropriate selection of chemicals and LODs in the left tail of the lognormal. Use of alternate procedures to handle LODs was also recommended. The European Food Safety Agency has published some useful guidelines in this area (EFSA, 2010). In addition a textbook, entitled "Nondetects and Data Analysis" was also recommended by the Panel (Helsel, 2005).

c. Not all parameters were measured during each sample collection cycle thus, potential for temporal trends exposures in some cases were not identified. Some consideration may need to be given to this.

d. The universe of chemicals targeted in NHANES was relatively small (approximately 200). In addition, the Panel noted that selection of chemicals for NHANES monitoring involves a lengthy interagency nomination process that may not target chemicals that have high potential for exposure. For

example, 17 sulfonyl urea herbicides are targeted in NHANES methods; yet use patterns suggest that potential for human exposure are small. There were no reported detections of these compounds in any NHANES samples (CDC, 2014).

In conclusion, the Panel acknowledged that many of the limitations of the SEEM approach were highlighted in the "Limitations Section" of the White Paper and in a recent journal article describing the Expocast project (Wambaugh et al., 2013). Thus, it appears that the Agency is well aware that progress is needed in many areas to reduce uncertainty of exposure estimates. Overall, the Panel agreed that SEEM's underlying science appears sound and represents a significant step forward in HTE assessment. It was also recognized that the process is on-going and model developments are at varying stages of maturity. The Panel unanimously commended the EPA scientists for the quality of their presentations displayed during the public meeting. In many cases, the presentations provided clarity that was lacking in the Agency's White Paper.

**Charge Question 1.2 Please suggest the most important steps for EPA that could decrease uncertainties and increase confidence in using the HTE approach to predict exposures in various demographic groups (e.g., young children, women of child bearing age) for large numbers of chemicals.**

In general, the Panel commended the EPA for the steps taken to include exposure data in the prioritization process along with toxicological data. The Panel noted that the NHANES data on which SEEM heavily relied on for model calibration and generation of uncertainty estimates were limited in scale and scope. Discussions focused on several potentially significant limitations that were described in the Question 1.1 response.

The Panel recommended that EPA work closely with CDC to help determine which data are most needed from NHANES for model validation and calibration. Also, EPA should critically evaluate each biomarker used for validation and calibration to ensure it is the best biomarker. Toxicokinetic information should be included to account for the fraction of the biomarker present in urine.

The Panel noted that exposure estimates do not account for temporality or magnitude spikes. This is a concern since critical time points in pregnancy may be very dependent upon this temporal/magnitude element. The Panel suggested that temporal data collected during pregnancy to evaluate exposures during critical pregnancy time windows may also be used. The HTE/SEEM approach may be improved by considering other biomonitoring exposure data in addition to NHANES. The National Children's Study or US birth cohorts (Eskenazi et. al, 2003) are other valuable sources that the Agency should consider.

The current SEEM approach used total population samples to develop the 5-parameter (factor) model (Fig 17). The subgroup analysis was then performed based on those same five factors for each subgroup. The Panel suggested building the prediction model for each subgroup (e.g., young children,) separately, including prior probability assessment, Markov chain Monte Carlo (MCMC), and posterior distribution in model development and calibration. The Step-wise regressions, may be used, to identify best subset factors. Different subgroups would likely identify different best subset variables.
For the models used EPA, the Panel suggests that the Agency considers the following questions/factors:

- What differences in person-specific model parameter values might occur from the usual adult?"

- Is there any reason to restructure the model (change the model-based equations) for particular demographic groups?"
- Are physicochemical characteristics of chemicals unchanged for demographic groups?
- What is the direct exposure from consumer products?
- Are these exposure factors different for children?
- Are there other demographic group differences?
- Are exposures to chemical emissions into the environment different for different demographic groups considering exposure routes and pathways?

It was emphasized that it is important to distinguish HTE model parameters/inputs for women of child bearing age from those of the general adult population. Toxicokinetics (TK) should not be different, generally speaking. For young children, clearance capacity per kg Body Weight (BW) can be higher than for adults. This is generally true for renal Clearance (CL) with glomerular filtration rate (GFR) (mL/min/kg BW) peaking at approximately 30 months of age at greater than twice the rate of the usual adult value. However hepatic CL ontogeny is complicated. The Cytochromes P450 (CYP) enzymes involved in xenobiotic metabolism increase after birth to adult levels over different time periods. Hepatic Intrinsic Clearance (Clint) is similar to or lower than usual adult levels, although fraction unbound to the chemical (fup) tends to be elevated in children, which makes the CLh (Hepatic Clearance) closer to the usual adult value. For the systemic exposure concentration at the steady state (Css) area under the curve the curve (AUC)), young children may have less exposure than adults for the same intake rate of a chemical. The question of sensitivity differences, between young children and adults, to a particular systemic exposure, is much harder to answer and should be a potential focus area for EPA. Finally, the Panel expressed a desire to see examples of the validation and calibration in the White Paper which would make EPA's approach easier to understand and enable reviewers to see best case and worst case scenarios.

**Charge Question 2.1. In the absence of sufficient empirical toxicokinetics information for the thousands of EDSP relevant chemicals, please comment on the approach of using HTTK, HTPBTK, IVIVE, and RTK for estimating chemical TK to provide an administered dose context to the concentrations showing bioactivity in the endocrine-related HTS assays**.

The Panel commended the EPA scientists for approaching this difficult problem with scientific rigor. Panel members felt that the oral presentations were excellent and greatly assisted with interpretation of the materials. The main conclusions that came from the Panel were that the EPA is going in the right direction and that there were no other existing viable approaches. Furthermore, there is a large body of published peer-reviewed scientific literature that supports the Agency's approach. The main assumption that is the premise of this approach appears to be a valid assumption that the free concentration in water at the biological target (receptor) will be bioactive. Additionally, the use of a combination of *in vitro* and *in silico* approaches to estimate exposure is quite novel. Embracing uncertainty analysis is also highly commendable. The EPA is doing an excellent job and is setting up tools now that will be augmented if/when additional datasets become available. The model development itself will identify data gaps that will need to be filled. The techniques are widely and successfully used in the pharmaceutical industry, when corresponding *in vivo* data are available and there are limitations that impose constraints on application.

Recent advances in the following including: a) the *in vitro* prediction of hepatic clearance using hepatocytes and microsomes (Riley et al 2005), and of tissue blood distribution coefficient using tissue composition data and physicochemical properties (Schmitt, 2008), and b) the availability of HT

methodology for *in vitro* measurement of plasma protein binding and hepatic metabolism make available parameter values required for TK and PBTK models to be used for large numbers of chemicals. Thus, it becomes feasible to use HT methodology to rapidly determine a threshold critical concentration of chemical at sites of toxicity and convert this concentration to an associated dose rate. This approach will require chemical specific information including the chemical plasma protein binding (free fraction in plasma at the bioactive concentration), the intrinsic hepatic clearance, and for PBTK methods, the tissue: blood distribution coefficients. However, the distribution coefficients may be unnecessary for calculation of a steady-state concentration (Css) as the volume of distribution of the tissue compartments has no role in its value.

However, as noted by the Agency and the Endocrine Policy Forum during the public meeting (and agreed upon by the Panel) there are limitations. The limitations includes those that surround measurement limitation (non-detects), availability of data, testing on just a subset of chemicals that does not represent the diversity of chemicals, the selection of variables and distributions for Monte Carlo simulations, adequate representation of bioavailability and absorption, using QSAR as a proxy which may fail to capture extra hepatic metabolism, secretion in bile/enterohepatic recirculation, unusual binding affinity for one tissue, substrates for transporters in renal, hepatic, and other tissues, and poor absorption. The Panel highlights other assumptions and limitations that are detailed in question 2.2, 2.3, 2.4 and 2.5.

**Charge Question 2.2. A comparison of the HTTK-predicted steady-state blood concentrations with *in vivo* values from the literature suggests that the overall correlation is low, but that the discrepancy between the two can be predicted using a combination of chemical properties, quantitative structure activity relationships (QSAR), and cutoffs from the HTTK assays. Please comment on: a) how well this approach characterizes the uncertainty in the steady-state blood concentrations, and b) whether the identification of chemical classes that need additional TK investigation is useful in a chemical prioritization or initial screening context.**

General Comments
Figure 28 page 78 of the Agency White Paper depicted the "Comparison of $C_{ss}$ predicted from *in vitro* HTTK with *in vivo* $C_{ss}$ values determined from the literature". This question is focused on the relationship demonstrated in Fig. 28. The Panel agreed that Fig. 28 clearly demonstrated a poor correlation between HTTK-predicted Css values and *in vivo* Css values from the literature. The White Paper then went on to describe how compounds with high residuals may be identified using a recursive partitioning algorithm (Fig 29, page 79). The Panel agreed that this is an appropriate approach.

The Panel recommended that a detailed analysis of Fig. 28 is necessary to determine whether or not there really is a defendable positive relationship. While the data set used to generate the figure was relatively small, a higher correlation would reasonably be expected. If the data within Fig 28 are considered alone, i.e. with only the data points shown without the residual bars and without the diagonal line, the impression is that there is minimal significant trend between the predicted and experimental data. The (visual) impression of a positive slope depends on a small number of extreme points. Another point to note is that the range of the data in Fig 28 covers several orders of magnitude on both axes; and it appears that the model has no predictive value within one order of magnitude. Therefore, in order to demonstrate the validity of this relationship (i.e. that there is a significant model underlying this prediction) and hence for predicted values of Css to be truly useful, *in vivo* data for further chemicals should be added to assist in the calibration.

Related to the comments on the relationship shown in Fig. 28, the White Paper did not discuss the quality of the *in vivo* Css values, and provided little analysis or discussion as to why the correlation was so poor. It seems that effort should be focused on understanding the failure of the HTTK model to better predict the *in vivo* Css. The White Paper (p. 85) listed several possible model deficiencies that might contribute to the poor correlation (extrahepatic metabolism, enterohepatic recirculation, unusual binding affinity for one tissue, transporters, poor absorption), but it seems unlikely that these deficiencies would apply to most of the chemicals in the test data set, although most of the chemicals were not identified. Without a better prediction performance, the HTTK predictions will not garner much credibility. Fortunately Figs. 31 and 32 of the Agency's White Paper demonstrate that the HTPBTK model predictions give considerably better correlations with the corresponding *in vivo* metrics, which suggests that HTTK has the potential to provide more accurate predictions of measured Css values.

*a) How well this approach characterizes the uncertainty in the steady-state blood concentrations,*

There is a great benefit in predicting steady-state blood concentration and this is vital part of predicting exposure within ExpoCast. As described in Chapter 3 of the White Paper, the EPA has made significant progress in predicting toxicokinetics, and Css in particular. The process of using TK, and the role of prediction is rational and in conjunction with much of the thinking in the European Union (EU) (e.g. within the SEURAT-1 initiative (http://www.seurat-1.eu/) and other projects e.g. ChemScreen http://www.chemscreen.eu/) and Predict-IV (http://www.predict-iv.toxi.uni-wuerzburg.de/en/) as well as the on-going HESI Risk21 initiative (Embry et al 2014). The finding that data from a small number of *in vitro* assays and QSAR descriptors could potentially provide a means of determining, for some molecules at least, Css is a significant step forward. Key to understanding and using the modelling approach is the characterization of the model(s) and data on which they are based.

The approach described in the White Paper made good progress in identifying uncertainties in the Css and in providing methods to characterize them. This is important for many reasons in particularly to: a) understand the poor relationship shown in Fig 28 and b) provide input parameters for the recursive partitioning tree shown in Fig 29.

There may be a case to consider further the factors that contribute to the current uncertainty in prediction of Css. The issue of *in vivo* data quality is noted above. In addition the current *in vitro* data consider TK processes in general terms and there may be a case to consider individual processes (e.g. for metabolism, see Agency Charge Question 2.2b).

Another source of uncertainty is the chemicals themselves changing the biological properties of binding proteins/receptors/ kinetics. This could be a source of error for *in vitro/ in vivo* comparison (epigenetics, affinity/capacity) - and affect fraction unbound and model assumptions required for HTTK and HTPBTK models. The Monte Carlo simulations that simulate biological variability are designed for individual variability and may potentially be used to accommodate for this source of uncertainty. The suggestion by the Endocrine Policy Forum about using Monte Carlo techniques to address exposure variation related to evaluating HTS experimental parameters that may affect HTE-HTTK-HTS correlations is related to this comment.

The use of a recursive partitioning approach is appropriate. However, only a small number of descriptors were considered. It may be appropriate to expand this number from the current list of 18 descriptors to include other relevant parameters (e.g. stability, functional groups, binding to Glutathione (GSH) etc; the toxprint list may be of use (www.toxprint.org)). However, it should be remembered that the recursive partitioning approach described in Fig. 29 can be unstable when the number of variables is

large. Overall, recursive partitioning is a good technique to analyze the residuals. This is a transparent technique that allows for rules to be created, but cross-validation to validate the resulting model is important.

The regression tree divides the samples into disjoint subgroups according to their QSAR descriptors so that chemicals in the same subgroups would have similar descriptors. This approach can be useful if each chemical group can be characterized in terms of QSAR descriptors as well as other chemical characteristics/properties. Further work may be considered on the cross validation of the regression tree in the context of it making predictions. For example, the 82 chemicals may be split into two groups and the same analysis re-performed; that is, use one group to build the regression tree and another group for residual analysis. This process may be repeated many times for different two group partitions.

The Agency's concept that some chemicals do not reach Css, or it takes many years may appear logical. However, it would be helpful if evidence could be provided to support this hypothesis. If proven, this could be useful to the EPA to demonstrate the validity of the approaches and general predictive model.

A specific comment relating to p. 74 and Fig. 26 (page 75) in the White Paper was linked to the comparisons of peak concentrations with Css predicted for an infusion scenario – when a chemical is rapidly cleared and there is no accumulation with repeated dosing, the predicted Css overestimates the Cpeak (peak concentration). This seems counter-intuitive, that the predicted Css would be very low and that the Cpeak would be larger than the Css, not smaller.

The Agency specifically raised the issue of whether the predictive approach could be for prioritization and / or screening. There was no consensus on this point from the Panel. At least one member held the view that the predictions for Css provided some possible utility for screening and prioritization, if used cautiously. However, others on the Panel held the opinion that the models were not yet suitable for this purpose and required further development. Some of this discrepancy may be because of the lack of clarity surrounding what was presented in the White Paper and what was presented during the meeting (please refer to the response to Charge Question 2.5).

In response to EPA's question on the limitations of the approach leading to uncertainty, the Panel identified following:
- Lack of definition and/or characterization of the underlying models (i.e. that it is more difficult to interpret the findings).
- Domains of models and *in vitro* systems not being stated, therefore, it is not possible to determine if the poor predictions may be classified as "reliable" i.e. scientifically valid and within the domain.
- The quality and accuracy of (and hence error associated with) the data modeled and for which predictions have been made are not stated.
- The possible uncertainties noted by the EPA e.g. extrahepatic metabolism, enterohepatic recirculation, unusual binding affinity for one tissue, transporters, poor absorption are not fully evaluated and this evaluation would be useful to do.

*b) whether the identification of chemical classes that need additional TK investigation is useful in a chemical prioritization or initial screening context.*

The Panel agreed that identifying chemical classes that need additional TK investigation is useful. Implicit in this statement, although not stated, is that chemical classes that do not require additional

information for which to make a decision for prioritization / screening could also be identified, this in itself is a significant step forward in model development and the application of the models. The concept that a single model does not fit all is often overlooked.

Chemicals of concern in the environment are structurally highly diverse. From Fig. 28 it appears that HTTK will more accurately predict Css for some chemicals than for others. A chemical class system that would identify the poorly predicted chemicals in advance would be useful. In Fig. 29 chemical class descriptors were identified (Fub, pKa, log P, MW) that seem to affect the ability of the HTTK model to accurately predict Css. Further experience using the HTTK approach may identify other relevant classifiers, e.g., water solubility, number of H-bonds formed with water, and structural features associated with particular transporters.

If chemical classes are to be used they need to be defined sufficiently and fully. This, to some extent, relies on what is meant by chemical classes (i.e. their definition in terms of whether they are analogues, share a common function group, mode of action, metabolic pathway etc). The Agency could benefit by collaboratively addressing this with the EU for Regulation on Registration, Evaluation, the Authorization and Restriction of Chemicals (REACH)) and at the Organisation for Economic Co-operation and Development (OECD). One of the key problems is the lack confirmation regarding how chemicals are membership classed. This is important as chemical classes can (if not defined strictly) be very broad. It is well known that very small changes in chemical structure, as minor as changing an atom or position of a functional group, can have very significant effects on metabolism, uptake etc. They may also affect ionization which is important for distribution, etc.

Particular chemical class(es) that may be of concern are the pharmaceuticals and possibly other "active" compounds such as pesticides. These compounds and their associated classes are not only designed for certain use and properties, but they are also designed to be bioavailable. The EPA should consider whether data for such compounds, and hence models derived or evaluated from such data are applicable to non-specific chemicals in particular including some of the Personal Care Products which are designed to be specifically non-bioavailable. The same argument may be made for many of the *in vitro* /*in silico* models for TK which come out of the pharmaceutical industry and may be developed and optimized for pharmaceuticals e.g. SymCyp but also other approaches.

Finally it may be better in the long term to identify the reasons why a particular chemical classes are poorly predicted and even consider these as a means of "grouping" to identify compounds for which further TK information may be required. The formation of these groups may be cyclical as more information is gained and fed back into the modelling and evaluation process, etc. For instance, there may be specific routes of metabolism which are not well modelled. Some approaches are available (or may need to be developed) to identify specific routes of metabolism. Such computational approaches were reviewed recently by Kirchmair et al (2012).

**Charge Question 2.3**. **Please comment on whether the assumptions made in these models are appropriate given the current state of the science and data limitations.**

The HTTK model was a compartmental model with assumed first-order oral absorption ($ka = 1/h$) and elimination by glomerular filtration and hepatic metabolism. The HTPBTK model was relatively more complex with compartments specified for organs of chemical uptake and elimination and the remaining tissues lumped into a "rest of body" compartment. Both models were suitable for their intended purposes. The HTPBTK model may be modified readily to explicitly include any tissue of particular interest (i.e., a tissue that was the site of toxicity). The models had several assumptions in common,

which are listed in the following table; if an assumption is specific for only one model that is indicated in Table 1 below.

| Table 1. HTTK Explanation of Model assumptions | |
|---|---|
| *ASSUMPTION* | *ASSUMPTION APPROPRIATENESS* |
| **Absorption ka = 1 h$^{-1}$** | This assumption is common and appropriate for calculation of Css and AUC values which were not affected by absorption rate. When a peak plasma concentration after the repeated oral administration of impulse doses is calculated, it is sensitive to the ka value, particularly for chemicals where the depuration t$_{1/2}$ was less than the time interval between doses (8 hr.). Considerable person-to-person variability has been observed generally in the oral-route ka value for different chemicals Therefore, a ka value of 1 h$^{-1}$ represented a typical value and a reasonable assumption. |
| **GI bioavailability = 100%** | This assumption likely overstated the fraction of dose absorbed from the GI tract for many chemicals; however, the model accounted for hepatic first-pass elimination. Overstatement of GI bioavailability would lead to overestimation of blood concentration metrics for incompletely absorbed chemicals. The assumption was appropriately conservative, given data limitations and absent the use of a QSAR model for GI bioavailability. |
| **Model Structure** | The PBTK model represented the body as compartments for the gut, liver, lungs, kidneys, and blood, and lumping of tissues to a single compartment termed "rest". Separate compartments for portals of entry and exit, and all other tissues lumped together as "rest", seemed appropriate for the intended purpose of the model. Specification of additional tissue compartments would be possible if they were needed. The TK model further lumped the lungs, and kidneys into the rest-of-body compartment, which was appropriate for estimation of steady-state blood concentration. |
| **Flow-limited Distribution in Tissues** | This assumption was generally appropriate.  However, the assumption was not appropriate for chemicals that pass slowly from blood to tissue space (polar and high molecular weight (MW), except for simulation of steady state.  For permeability-limited distribution, time to steady state would be underestimated. |
| **Instantaneous Distribution within a Compartment** | Each compartment was assumed to be well mixed; i.e. There were no concentration gradients within a compartment. In general, this is an appropriate and accepted assumption particularly since the word "compartment" is used to help define define/explanation the term Toxicokinetics (TK)." |
| **Linear (concentration-independent) Distribution in Tissues/Compartments** | The concentration within a compartment was assumed to be proportional to the blood concentration when the system was at steady state. This was appropriate and has been commonly assumed by TK modelers. |
| **Vd for Tissues** | In the PBTK model the volumes of the tissue spaces (via tissue/blood distribution coefficients) were calculated using the approach of Schmitt (ref. 89).  Schmitt's validation results indicated that the approach was quite accurate and therefore it was an appropriate approach to HT estimation of Vd values.<br><br>It was unclear how the distribution aspect of the model would work for chemicals that do not distribute into intracellular space, and how such chemicals would be identified.  Such chemicals generally may not get into the body via the oral route, absent an active transport system, but they may enter via broken skin or inhalation. |
| **CLint from *In Vitro* Measure** | This methodology seemed to work fairly well (Riley ref 100). Observed vs. predicted CL values showed good correlation and generally agreed within an order of magnitude. This degree of agreement was acceptable, considering the large person-to-person variability that has been observed for this parameter. |
| **Linear (concentration-independent) Hepatic Clearance** | Appropriate (probably) given the typically small dose rate.  Saturation of hepatic CL *in vivo* for drugs has been rarely observed and only at relatively large doses. |
| **CLr = fup x GFR** | With this assumption renal tubular reabsorption and secretion were ignored and for many chemicals this would overestimate CLr due to reabsorption or underestimate CLr due to active tubular secretion.  This would lead to under- or over-estimation of blood concentration metrics. |
| **Intake via GI Tract** | Dermal and pulmonary routes were ignored, but *in vivo* Css values were calculated for oral intake of 1 mg/kg/day, so the assumption was appropriate for this situation. The HTPBTK model had provision for pulmonary uptake of inhaled chemicals, and it was readily modifiable to include dermal absorption. |
| **Metabolites ignored** | Metabolites could be important contributors to toxicity and modulators of parent chemical TK. |
| **fup Default value** | Random draws were selected from a uniform distribution from 0% to 1%. This approach appeared superior to simply using a value of 0.5%. A small fup value has been generally associated with relatively large, person-to-person variability; e.g., exceeding a 10-fold range in the case of warfarin, which supported the use of a broad range of values between 0% and 1%. |
| **Assumed Distributions for MC** | A log-normal distribution was generally used; exploration of other distributions may be informative and build confidence in the MC simulations. |

The Panel advised that a sensitivity analysis of TK model parameters may be helpful to assess whether assumptions were appropriate. If model predictions were relatively insensitive to changes in the value of a parameter, then assumptions associated with that parameter would be less likely to affect the predictions of the model when the assumptions were not exactly correct. Attention could then be focused on assumptions associated with model parameters that produce the greatest sensitivity in model predictions.

**Charge Question 2.4. Please suggest the most important steps EPA should take to improve the various kinetic models to provide rapid and cost-effective predictions for large number of chemicals.**

To improve the various kinetic models, the Panel recommended that for the HTTK and HTPBTK models, there should be an attempt to find the cause(s) of the poor correlation with Css, literature as specified in Fig. 28 of the Agency White Paper. The HTTK model should have given predictions that were closer to those calculated from known PK parameters. Either the literature values were incorrect (seems unlikely) or there were model deficiencies that might be correctable.

One possibility was the assumption that CLr = fup x GFR. Log P values may provide a way to introduce tubular reabsorption, which has the potential to markedly reduce renal clearance. One question to explore is whether or not there are QSAR or HT screens for renal tubule transporters. Most drugs are not cleared extensively by the kidneys when there is low fup and/or tubular reabsorption.

Another possibility was oral absorption, which may have been less than the assumed 100%. The Panel recommended that a QSAR model be developed to predict the oral absorption fraction. Approaches have been reported in the pharmaceutical literature for drugs (e.g., Lipinski, refs. 106 and 107). Also HT screening methods have been described; the PAMPA system characterized chemical membrane permeability using a 96-well-plate platform that was capable of a throughput of 500 – 1000 chemicals per day (M. Kansy, et al., 1998). In this regard, dermal permeability coefficient estimation relationships that use Kow and molecular weight (MW) were described in ref. 122, p.60 of the White Paper.

In regard to other important steps, the Panel recommended refining the HT *in vitro* estimation of hepatic clearance to enable a lower concentration range of determination. On p. 76 of the Agency's White Paper it was acknowledged that the current use of a hepatocyte suspension was insensitive to measurement of CLint for low turnover compounds. Development of more sensitive analytical methods may be an approach.

Also, the Panel suggested refining the HT determination of fup, to improve accuracy in the low concentration range. As described on p.88 of the Agency White Paper, the use of diluted plasma in the RED assay and extrapolation of fup to undiluted plasma may help. Spectroscopic methods have been described that may prove suitable; e.g. nanospray desorption electrospray ionization mass spectrometry or matrix-assisted laser desorption/ionization. Development of more sensitive analytical methods may also be an approach.

For the HTTK and HTPBTK models the Panel also recommended extending them to include dermal and inhalation absorption pathways. In addition, the use of additional animal data (other than the taxa reported in the White Paper) for TK model validation may expand the number of chemicals and build confidence in the TK model prediction capabilities.

The Panel advises that all *in silico* models for kinetics (e.g. PBPK and the tissue partitioning models) would benefit from better characterization and description than what was provided in the White Paper. This would give confidence to users of the models by improving their transparency and allowing the users to see how the predictions are being made. In addition, as significant improvements and new models are developed, this would promote easier use of the models could be used more easily by others, this would provide a welcome advance in the science. In particular, where possible, the equations or algorithms should be stated. This may be supported by performance statistics such as goodness of fit, robustness, predictivity etc. If possible, the applicability domains of the models may be provided which may include information on the chemical space (i.e. properties or structural fragments etc). Estimation of which of the EDSP chemicals are within the chemical space of the models would give greater confidence in the prediction of their properties. The Agency should consider basing he description of the models could be based around the OECD Principles for the Validation of QSARs (see http://www.oecd.org/chemicalsafety/risk-assessment/validationofqsarmodels.htm).

As well as full documentation of the kinetics models, the Panel encouraged the EPA to make them publicly available and to publish them in peer reviewed journals. Release of the algorithms developed in R programming language would be ideal and very helpful to the greater scientific community. A guide on how to use these models would be extremely helpful.

Currently, the EPA presented exposure modeling by the oral and inhalation routes. It is recommended that the EPA also consider the dermal route of exposure in PBPK modeling. This will be useful for other exposure scenarios (e.g. personal products). However, there are fewer PBPK models for dermal exposure. Work being undertaken by the EU COSMOS Project (www.cosmostox.eu) may be useful in this regard.

The White Paper described the development of linked models that incorporate information from both *in vitro* assays and *in silico* models. The EPA should consider use of the KNIME technology as a means of combining together chemoinformatic tools. KNIME is a freely available workflow platform (www.knime.org). R (program language) models can be combined into the workflows and may provide an ideal platform for dissemination of these models.

It appears that within some of the models there is a reliance on predicting pKa. This is still problematic for more complex chemicals and consideration may be required as to when this is appropriate (See Panel responses to question 1.1).

**Charge Question 2.5. Please comment on the approaches presented in the present White Paper for comparing RTK-derived endocrine-active dose ranges with exposure predictions from ExpoCast. Discuss the strengths and limitations of these comparisons, and whether this or other approaches are suitable for clearly distinguishing chemicals with higher predicted doses from chemicals with lower predicted RTK adjusted doses.**

*Note:* There was some initial confusion surrounding Charge Question 2.5 because the appropriate figure comparing RTK-derived endocrine-active dose ranges with exposure predictions was not included in the SAP White Paper. The appropriate figure, that summarizes the findings, was actually included in the Agency's PowerPoint presentations during the public meeting. In the following, the *first part* of the response to Charge Question focuses specifically on these comparisons and on "whether this or other approaches are suitable for clearly distinguishing chemicals with higher predicted doses from chemicals with lower predicted RTK adjusted doses," as per the Charge Question. In a *second part* of the response

the Discussants provide more general comments pertaining to the methods that produced the ("reverse" and "forward") estimates and predictions related to this Question.

*Panel Response Part I.*
As noted, the figure that Question 2.5 is specifically referred to, was presented by the Agency at the public SAP meeting (slide 18 of presentation titled "High throughput Toxicokinetics (HTTK) and Reverse Toxicokinetics (RTK) for EDSP"). This figure plotted individual compounds against RTK oral equivalents (using data from a suite of assays). When this plot was compared to the raw ToxCast AC50 values versus compound figure (slide 17 of the above mentioned presentation), there was no pattern, however it can be stated that now there is clearly a pattern distinguishing compounds that have higher RTK oral equivalents and those with lower oral equivalents. The figure that plotted compounds versus the RTK oral equivalents was created using a three-compartment model that estimated steady state concentrations in plasma using plasma binding and hepatocyte clearance assays and assuming linear relationships and steady states for daily dose and oral equivalent in *in vitro/in vivo* extrapolations. When these assumptions are acceptable (for the compounds and concentration ranges considered) and uncertainties are taken into consideration (previous Charge Questions dealt with assumptions and their limitations), then this method should be able to distinguish the higher predicted dose range from the lower predicted dose range (although middle-range doses may not be clearly distinguished from each other). The Panel was not aware of any alternative approach capable of providing better results than this method given the (minimal) amount of information available for this type of analysis. However, caution was advised: the comparison of Css estimated by these *in vitro* methods and compared to published *in vivo* Css values did not show a strong relationship (an issue covered in the Response to Charge Question 2.2). However, results from the HTPBTK modeling, applied to a few chemicals, suggest that the Css values estimated from three-compartment models appeared to do an adequate job of reproducing peak PBPK concentrations derived from the HTPBTK models (although the comparison is performed on a logarithmic scale) and that the parameters estimated from HTPBTK (AUC, CMax) seem to correlate reasonably (or are conservative) with *in vivo* parameters from the literature. This appears to be a good start, but the process has only been completed for a few chemicals; therefore, further evaluation and validation with a wider range of chemical classes are needed.

Finally, the Panel noted, as mentioned in Charge Question 2.5 the description of the Agency's approach should be better clarified and simplified. The Panel stated that it was difficult for them to decipher the Agency's methods and they had to search and to repeatedly read the White Paper to understand the exact methods used by the Agency.

*Panel Response Part II.*
The following comments summarize a more general discussion pertaining to the issues raised by Charge Question 2.5 and address related aspects of the following objective stated on page 64 of the Agency's White Paper: "We (the Agency) attempt to compare exposure predictions to exposures inferred from monitoring data, and we do so systematically across as many chemicals as possible to quantify the impact of these assumptions."

Given the approximate nature of HT screening/ranking, the evolving approaches presented in the White Paper for comparing RTK-derived endocrine-active dose ranges with exposure predictions from ExpoCast should be considered generally appropriate, taking into account that uncertainties in both "forward" and "reverse" (RTK-derived) intake estimates span multiple orders of magnitude. The limitations of both forward and reverse predictions (as well as of the data from NHANES that were used for calibrations) were discussed in previous questions and should also provide the context for the answer to the present question. Therefore, comparisons in principle should be limited to situations where

exposures are low and follow regular repetitive patterns that would support the assumptions (linearity, steady state) incorporated in the HT models. [1]

The Panel noted the following positive aspects regarding the Agency's approach:

(a) there is substantial methodological improvement with each "generation" of SEEM analysis,

(b) the limitations of heuristic HT approaches appear to be recognized and the context of HT screening is clarified with the progress of the analyses, and

(c) a rational "path forward," that incorporates the critically important near-field chemical "releases," and involves the refinement of a more mechanistic approach utilizing SHEDS-HT, appears to be solidly in place.

It should be emphasized, however, that:

- It is necessary to define precisely the appropriate exposure metrics, either individual-based or population-based, that are to be compared after they are produced by different, forward and reverse, modeling exercises. (Otherwise there is a distinct danger of comparing "apples to oranges").

- In comparing model predictions with observations it is important to list correspondence (temporal, demographic, etc.) of the attributes associated with each and to match these attributes, to the extent possible, in order for the comparisons to be useful.

- Distributional comparisons (that incorporate explicit estimates of variability in the metrics of exposure) should be the target for future work.

- Comparisons should eventually go beyond the NHANES data and focus on sensitive subpopulations of concern for cases where relevant data are available.

- Systematic global sensitivity/uncertainty analysis should accompany future SHEDS-HT applications to characterize individual contributions.

In the first generation SEEM analysis, which considered 1936 chemicals with exposure estimates developed using far-field mass balance models (USETOX and RAIDAR) and a yes/no indicator for near-field "chemical use,", a steady-state RTK approach was used to infer exposure for 82 chemicals from NHANES biomonitoring data (creatinine adjusted urine concentrations). (It is mentioned, however, on p. 47, that 1850 chemicals without NHANES data were considered, which would make the total 1932 instead of 1936). It was assumed that that the individuals were at steady-state due to a constant rate of exposure; an average daily creatinine excretion and an average daily urine volume of 1.4 L were used along with a mapping between parent and metabolite compounds to convert urine concentration to an exposure in units of mg/kg/day.

The models for HTE estimation (USETox and RAIDAR) used estimates of key aspects of exposure space: (a) physicochemical properties, (b) product characteristics, (c) emission characteristics, and (d) exposure pathways. They used fugacity and degradation properties and assumption of steady state to predict chemical fate and distribution in environmental media (air, water, soil, sediment, biota). The models then predicted an overall population intake fraction, kg absorbed per kg emitted. The intake

---

[1] A systematic discussion of the limitations of "reverse" (RTK-derived and RTBPTK-derived) exposure/intake modeling from biomonitoring data (especially in the context of real-world, non-linear, non-steady exposures) can be found in:

- Georgopoulos PG, Sasso AF, Isukapalli SS, Lioy PJ, Vallero DA, Okino M, Reiter L (2009) Reconstructing population exposures to environmental chemicals from biomarkers: Challenges and opportunities. *Journal of Exposure Science and Environmental Epidemiology* 19:149-171. doi:10.1038/jes.2008.9

fraction predictions of USETox and RAIDAR correlated well for "bioaccumulators", but not for compounds with Kow < 1.

Comparisons of estimates from "reverse" and "forward" modeling employed a Bayesian methodology and linear regression (p. 40 of Agency White Paper) was used to compare RTK-derived doses with HTE inferred exposures. The regression results and their uncertainties were then extrapolated to chemicals for which there were no biomonitoring data.

Steady state was assumed, both for USETox and RAIDAR predictions of environmental media and exposure inference from the NHANES data. This seems a reasonable first approximation, but in the "real world" intermittent and sporadic exposures are often the case. Assuming steady state to model these exposures would probably underestimate the peak exposures, by quite a large degree if exposures were of short duration with long periods of no exposure.

Chemicals with high intake fractions were somewhat underrepresented in the NHANES data, which decreased the confidence in the model predictions for such chemicals, a group that one would imagine contained a relatively high proportion of chemicals of immediate interest. The NHANES chemicals showed "reasonable coverage of the exposure space"; i.e., they were chemically diverse and showed several orders of magnitude range of intake fraction values with both HTE estimators. This strengthens the comparisons of course as it was desirable not to use the models on chemicals outside the exposure space; i.e., not to extrapolate beyond the training set of chemicals.

The first-generation SEEM analysis produced a rank-ordered list of chemicals from highest to lowest mg/kg/day doses (Fig. 14), with uncertainties that spanned several orders of magnitude. This seemed to mean that there were large groups of chemicals, the members of which were indistinguishable. The approach allowed identification of two distinct exposure groups – high and low – with an intermediate group situated in-between.

The second generation SEEM analysis used subject-specific information that included urine flow data reported by NHANES along with chemical monitoring data for 106 chemicals. This information was used to determine distributions of inferred exposure appropriate for ten demographic groups (defined by age, gender, and BMI classification). The urine flow rate and the subject's body weight reported by NHANES were taken into account, along with the parent and metabolite mapping, to convert urine concentration (µg/L) to an exposure (actually intake) in units of mg/kg bodyweight/day.
It was stated (p. 38 of the White Paper) that "Bayesian methodology was a natural choice to allow use of estimates that fell below the limit of detection. It was assumed that the population distribution of compounds in urine was lognormal." The exposures derived through RTK for the 106 NHANES chemicals were used to identify an "optimal set" of five chemical-specific descriptors that were used through a linear regression model, to rank 7,785 chemicals. The "high-throughput heuristic model was able to explain roughly half of the variance of the inferred exposures" (p. 61). The discussion of the analysis was appropriately cautious: On p. 58 it is stated "Although the heuristics seem plausible in retrospect, additional data may diminish or eliminate their significance. Similarly, plausible heuristics not supported by this analysis (e.g., vapor pressure) may be significant for chemicals not included in the current set of monitoring data. Although quantitative predictions are useful for prioritization, the statistical analysis here may be even more useful as a means to organize and prioritize the types of data needed for predicting human exposure to the thousands of chemicals in our environment. For chemicals that are similar to chemicals known to have high measured concentrations in the urine of the U.S. population, we need additional exposure information, such as biomonitoring and environmental

sampling. New biomonitoring data will also be needed for the chemicals whose properties are most different from those currently being studied in the NHANES and by the U.S. EPA."

The second-generation SEEM analysis used a higher number of test compounds in the training set, and identified five HT exposure heuristics. When the calibrated model was used to estimate exposure rates for 7,785 chemicals, the rank-order of the exposure rates showed a better separation of the chemicals than did the first-generation (Figure 19 compared to Figure 14). It therefore appeared that the evolving HT approach had improved potential to rapidly screen large numbers of chemicals and to distinguish them on the basis of their estimated exposure rate.

The approach taken in the third generation SEEM analysis was more mechanistic: it utilized SHEDS-HD, and it appeared that it aimed to explicitly account for variability as well as uncertainty in upcoming applications. The report presented (pp. 109 and 110) "SHEDS-HT exposure dose predictions for the U.S. general population for 2111 chemicals that were generated for simulated populations of 10,000 individuals. The range of the predicted exposures for these chemicals spanned 4 to 11 orders of magnitude, due to significant impact of high variability and uncertainty in the model inputs and parameters, but the results indicated also that a good discrimination among chemicals span for the purpose of prioritization remains feasible." Comparison of third generation analysis showed (p. 111) that the "median SHEDS exposure doses were significantly correlated with predicted median oral equivalent doses" and that "overall, the SHEDS distributions were higher than the NHANES values; this overestimation of exposure was not unexpected, as the screening-level estimates from SHEDS-HT were conservative by design."

*Overarching comment – Applying to Question 2.5 and to the White Paper in general*
The USEPA should be commended for the substantial effort and the progress in pursuing the work described in the White Paper. There is a need however for a more thorough editing of the document – ideally rewriting certain parts for clarity, but at a minimum correcting inconsistencies and improving the clarity of certain statements. The need to balance scientific rigor versus the requirements of High Throughput modeling makes certain compromises unavoidable – but the limitations they impose on the analyses should always be explicitly stated. It would be very useful to summarize, possibly in an overview table in the beginning of the document, the "evolution path" of SEEM stages, identifying the changes/improvements along this path (assumptions employed, major results, etc.).

**Charge Question 3.1. Please comment on the three key areas including whether there are other areas that are of equal or higher priority to support an integrated activity/hazard and exposure based prioritization and screening approach within the EDSP or other chemical programs.**

The key areas that the Agency should to take into account in order to enhance their efforts were identified by the Panel. The Panel first advised the Agency to reconsider its use of the terms "activity/hazard". The Panel noted the Agency's current usage of these terms may potentially be misinterpreted. Because the term "activity" is used extensively in exposure science in relation to human behavior, it would be useful to call the activity of a chemical "bioactivity". The term bioactivity provides a biological context and consequently avoids confusion.

Secondly concerning future efforts/key areas of focus, the Panel indicated that it is important for the Agency to recognize that exposure has multiple "dimensions". Thus, the metrics of exposure need to be defined and calculated with appropriate resolution (frequency) over appropriate time periods or "windows" (both with respect to "absolute" or "environmental" time and with respect to receptor

development and aging – e.g. exposures during pregnancy etc.). These metrics should be eventually relevant to potential health endpoints.
Examples of "generic" exposure metrics that should be informative in a High Throughput context are:
- Pervasiveness – a measure of the magnitude of potential exposures within the general population or within sub-populations of concern.
- Persistence – a measure of the temporal frequency and/or duration of exposures
- Severity – measure of the prevalence of high levels of exposures.
- Efficacy – a characterization of the contact a chemical which potentially results in intake, and to biologically relevant uptake, via one or more exposure routes.

Lastly, the Panel recommended that the Agency consider a plan for extensive use of a probabilistic HT exposure model, which consist of a basic "mechanistic foundation," (i.e. SHEDS-HT) within a framework that explicitly discriminates variabilities and uncertainties (e.g. via 2D-dimensional Monte Carlo analyses). This framework should address the development of three of the above mentioned exposure metrics including pervasiveness, persistence, and severity. Developing estimates of efficacy would require enhancements in supporting databases (e.g. CPCPdb, CPCat) to incorporate information on concentrations/amounts and physicochemical attributes of a chemical present in a product (e.g. dissolved in a liquid that is used either as a liquid or as a spray; used as part of surface coating or dispersed in a solid matrix; constituent of a powder, cream, or lotion; etc.).

**Charge Question 3.2. We propose applying the ExpoCast framework to ecological exposures using aggregated water monitoring data to evaluate the predictions of environmental fate and transport models. Please identify other data, models or environmental media that may be of greater value in the initial model calibration and uncertainty analysis.**

The Panel concluded that the proposed models were a positive movement by the Agency to address ecological exposure. In particular, the coupling of SHEDs to EXAMS was considered to be a good first step. However, the Panel cautioned that using aggregated water sampling to represent potential average environmental exposure may underestimate exposures in areas prone to greater discharge and higher concentrations. While probabilistic estimates may be obtained using this method, a much simpler and more conservative approach would be to utilize "worst-case" scenarios as initial screens. This process has been recommended for screening "emerging compounds" in the State of California (Drews et al., 2013; Maruya et al., 2013). It may also be appropriate to separate chemicals based on use categories. For pesticide residues, these scenarios would include agricultural drainage areas during seasonal/runoff potential. Discussions by recent FIFRA SAP's regarding atrazine's environmental fate and transport using PRZM-EXAMS may provide useful insight.

It was noted that use of monitoring data to estimate exposure is a significant departure from previous Agency practices used to evaluate potential ecological effects concentrations (EEC) for pesticides risk assessment completed for product registration. In this case the Agency relies almost entirely on simulation modeling. The rational for using monitoring data (as described in presentations) is typically based on the fact that high spatial and temporal variability in conjunction with varied sampling designs often biased source identification and delineation. In addition effective use of monitoring data requires development and management of archiving systems like ACTor appropriate QA/QC metrics which accurately identify analytical limitations. This can be challenging and labor intensive. Another concern is how to treat datasets with large number of non-detects (LOD). This is commonly observed when trace analytes like pesticides, residues of personal care products and other emerging contaminants are monitored in surface waters. Useful guidance is provided by the European Food Safety Agency (EFSA,

2010) and in a textbook by Helsel (2005).

One panel member suggested that in order to identify a mechanism to assess monitoring data for human health for drinking water derived from recycled water, the Agency should explore the findings of a Science Advisory Panel convened by the State of California charged with ascertaining emerging contaminants for monitoring (Anderson et al. 2010; Anderson et al. 2012; Drewes et al. 2013; Maruya et al. 2013). In addition, this panel member observed that for domestically derived compounds (i.e. Personal Care Products; "inert" ingredients; etc.) the use of monitoring data from wastewater-dominated streams from the Southwest or California may provide worst case data for confirmation of exposure estimates. In California, data may be obtained from Waste Water Treatment Plans (WWTPs; Publically Owned Treatment Works (POTWs) for secondary/tertiary effluents and often receiving stream measurements for multiple agents are available from regional state regulatory (California State Water Resources Control Board) and monitoring programs (SWAMP in California) or other regional monitoring agencies (Southern California Coastal Water Regional Project; www.sccwrp.org). Many POTWs serve more than 10,500 person/plant in California and waterway flows are often 100% wastewater. Thus, this situation would represent a worst scenario.

With regard to linking ecological water borne exposures to biota body burdens, panel members commented positively that the KABAM model incorporated trophic components. It appears that simple "One-box" models may also be useful for sediment, biota and loading estimates (Maruya et al., 2013). Further efforts to estimate target organ tissue concentrations (gonad/brain) may be accomplished using PBTK models particularly for fish (see responses to Question 2.1). PBTK parameters developed by Hayton and Schultz, as well as the Nichols group at EPA ORD (Duluth laboratories) may be utilized for certain species such as salmonids to estimate target organ concentrations. The target organ can be compared to HTP bioassays which are conserved in vertebrates. Finally, use of case studies for model demonstrations was highly recommended.

The Panel emphasized that an alternative or companion to monitoring data are concentration estimates derived by simulation modeling. A core EPA approach for pesticide simulations is use of the field scale model PRZM coupled to the EXAMS model. Thus, the Panel indicated that it may be useful to couple PRZM-EXAMS to SHED-PBTK. Other models that were recommended for use at larger spatial scales were SWAT (Soil Water Assessment Tool) and APEX (Agricultural Policy Extender Model). They were recently used in the USDA Conservation Effects Assessment Program National Assessment for Cropland (CEAP, 2014). The purpose of the assessment was to estimate the environmental benefits and effects of conservation practices applied to cultivated cropland and cropland enrolled in long-term conserving cover (e.g. the Conservation Reserve Program). SWAT and APEX were used to model water quality impacts of agricultural practices by determining the loss of pesticides in streams and rivers across the United States. Assessments were made at the basin scale and cover much of the continental USA. Pesticide simulations may be useful for "ground trothing" monitoring data or in providing inputs to exposure models at regional scales. All reports are available on-line (CEAP, 2014).

Other suggestions for the Agency to consider included construction of simple agent based particle models (with chemical properties assigned to the "agents") designed to explore various contaminant transmission pathways using environment in a GIS framework. These can then be used in pattern-oriented modeling (POM) to determine likely routes of transmission cross referenced with aggregated water monitoring data.

A panel member described POM as a technique that guides the development of models which simulate multiple patterns observed in the field and systematically tests how well agent-based models reproduce

such patterns. In other words it finds the most parsimonious and well-fitted model version. POM studies use traditional methods of goodness of fit such as the sum of squares evaluation or ad hoc comparisons of fitting errors and variations, and also pattern-oriented modeling information criteria (Piou et al., 2009). POM is a relatively new technique that may have broad application to systems that are data-limited with pressing management and conservation problems.

The Panel noted that the approach of using a simple model and applying different behavior to explain the patterns has been proposed for disease outbreak by Murray (2009). Murray (2009) modeled a theoretical disease to explore density independent, density dependent, non-linear density dependent and constant infection pressure to demonstrate that different characteristics of disease transmission can lead to different model behavior. For contaminant exposure, one could start with a simple model, then alter the environment (through GIS) to simulate different modes of contaminant transmission, and determine which scenario or combination of scenarios best explains the aggregated water sample pattern (via POM). Certain assumptions will be required. For example, assuming that currents do not vary from year to year (they may vary seasonally, but each year they roughly stay the same). Stochastic elements can also be added to the model to simulate fluctuating abundances that roughly match the coefficient of variation (CV) of annual abundance of elements that vary from year to year, and thousands of simulations can be run to capture the range of possibilities (parameter space).

In summary, POM is often used for ecosystem scale questions approached "bottom-up" and best used where data availability is often sparse and incomplete (Grimm et al, 2005). Patterns characterize systems and are indicators of the structure and function of the system. The General understanding of how most ecosystems are structured is limited. This is also true for contaminant dispersal to evaluate of exposure. In POM the pattern is used as the output. The model is varied in an attempt to discern the structure and function of the system. Outputs can be compared using common statistical tests (Chi-square, ANOVA, Spatial Statistics, AIC, POMIC; Peck, 2004) to determine the most plausible scenarios.

**Charge Question 3.3. Please identify other relatively high-throughput sources of exposure-related information that should be included in these three key areas and suggest how this information would be used to help prioritize chemicals.**

The Panel identified the following areas that require further attention and ("distributional") characterization appropriate for HT modeling. They noted that these areas are critical in driving exposures as well as doses. These areas entail ascertaining:
 (a) human behavior/activities including consumption/usage of consumer products and
 (b) attributes of microenvironments in which the human activities take place (both residential and non-residential).

In order to address these areas the Panel recommended that the Agency consider the following. Firstly, building/improving distributions of human activities/behavior, appropriate for use with HT models, that account for variabilities associated with age, gender, urban and rural location, socioeconomic status, ethnic and cultural background, etc. is important for both "forward" exposure modeling and for proper interpretation of chemical biomonitoring data via "reverse" exposure modeling. The USEPA 2011 Exposure Factors Handbook already provides various useful distributions that can be used as a starting point for High Throughput distributional exposure characterizations.

Secondly, the Panel advised that the Agency's Consolidated Human Activity Database (CHAD) should continue to evolve and be enhanced with information from local studies. Additionally, data currently in CHAD can be mined to build distributions appropriate for HT applications. Various other extant sources

are available from Federal and State Agencies and from Professional Groups or Associations. As examples, the PRoTEGE modeling system uses:

- age-group specific indoor/outdoor time spent distributions for each of the nine climatic regions of the contiguous US, that were derived from CHAD
- consumer behavior distributions derived from the Consumer Expenditure Surveys (CEX) of the Bureau of Labor Statistics
- a non-parametric distribution for ventilation rates of non-residential buildings constructed according using the Commercial Buildings Energy Consumption Survey (CBECS) data

Thirdly, the Panel advised that mining social media is an area that should be considered as a means for extracting patterns of human behavior, especially involving new products. Links to a wide range of on-line resources on human behavior/activities and on attributes of exposure-relevant microenvironments can be found at the Environmental Bioinformatics Knowledge Base (ebKB) web site (www.ebkb.org).

Lastly, the Panel suggested that coordination with EU REACH efforts is necessary, especially with respect to parameters that are chemical-dependent (rather than human behavior dependent), to avoid duplication of efforts (especially in QSARs, HTPBTK, etc.).

**Charge Question 3.4. For the HTTK work going forward, please comment on additional studies that could be performed or approaches that could be taken to improve rapid and cost efficient predictions of TK for large numbers of chemicals.**

A general suggestion was for EPA to consider and evaluate the applicability and usefulness of both:
  (a) additional QSAR methods for developing estimates for ADME parameters where currently default values/assumptions are used, and
  (b) cell cultures (including novel organotypic cultures and microfluidic systems) that aim to simulate biological processes and cellular interactions and provide parameters in a manner more relevant to actual tissues and organs compared to single-cell cultures.

Sources of uncertainties in HTTK models are associated with multiple biological processes including transport, enterohepatic recirculation, and most importantly metabolism. In the White Paper it was assumed that QSAR methods were sufficient in capturing the differences in tissue-blood partition coefficients. However, whether QSAR methods are sufficient in capturing the other processes is in question. For example, IVIVE of hepatic metabolism requires age-dependent quantification of hepatic CYP enzymes. EPA should consider cell culture systems (e.g. hepatic cells for metabolism, Caco-2 cultures for transport) for high-throughput to reduce uncertainties in predicting IVIVE using QSARs alone by evaluating a panel of cell culture systems (e.g. stomach epithelial cells, gut epithelial cells, nephrons, etc.) which affect chemical ADME most critically. This could be a part of the future direction as a way to reduce uncertainties from QSAR modeling alone, and also as a way to incorporate potential demographic differences in the prediction.

A major resource for availability, progress and applications of a wide range of organotypic cultures is the European Union SEURAT-1 Program ("Towards the Replacement of *in vivo* Repeated Dose"): information, documents and links can be found online at http://www.seurat-1.eu/.

# REFERENCES

Anderson, P. Denslow, N., Drewes, J.E.(Chair), Olivieri, A., Schlenk, D. and Snyder, S. (2010). Monitoring Strategies for Chemicals of Emerging Concern (CECs) in Recycled Water : Recommendations of a Science Advisory Panel. State Water Resources Control Board. 220pp. Available on line at:
http://sccwrp.org/ResearchAreas/Contaminants/StatewideCECPrioritization.aspx.

CEAP, (2014). USDA Conservation Effects Program Cropland National Assessment.. Drewes, J.E., P. Anderson, N. Denslow, A. Olivieri, D.Schlenk, S.A. Snyder, and K.A. Maruya (2013) Designing monitoring programs for chemicals of emerging concern in potable reuse- what to include and what not to include? *Water Science and Technology* 67:433-439. Available on line at: [http://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/technical/nra/?cid=nrcs143_014144]

Drewes, J.E, Anderson, P. Denslow, N. Olivieri, A., Schlenk, D . and Snyder, S. (2012) Monitoring Strategies for Chemicals of Emerging Concern (CECs) in California's Aquatic Ecosystems: Recommendations of a Science Advisory Panel. State Water Resources Control Board. 214pp. Available on line at: http://sccwrp.org/ResearchAreas/Contaminants/StatewideCECPrioritization.aspx

Eskenazi, B, Bradman A., PhD, Gladstone, E., Jaramillo, S., MA, Birch K., and Holland, N. (2003) CHAMACOS, A Longitudinal Birth Cohort Study: Lessons from the Fields, Journal of Children's Health: Vol. 1, No. 1, pp. 3-27 EFSA (European Food Safety Authority). Management of left-censored data in dietary exposure assessment of chemical substances. *EFSA Journal* 2010; 8:1557, 1-96. doi:10.2903/j.efsa.2010.1557. Available on line at: http://www.efsa.europa.eu/en/efsajournal/pub/1557.htm.

Helsel, D.R. (2005) Nondetects and Data Analysis: Statistics for Censored Environmental Data. John Wiley & Sons, New York..

KNIME (2014) The Open Analytics Platform Available online at: Available on line at: http://en.wikipedia.org/wiki/KNIME

Grimm, V., Revilla, E., Berger, U. Jeltsch, F et al. 2005. Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science* 310: 987-991.

Maruya, Keith,  Anderson, P., Denslow, N.,  Drewes, J., Olivieri, A.,   Scott, G., Snyder, S., Schlenk, D. (2013) An Adaptive, Comprehensive Monitoring Strategy for Chemicals of Emerging Concern (CECs) in California's Aquatic Ecosystems. *Integrated Environmental Assessment and Management* 10:69-77.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. *ASABE* 50(3): 885-900.

Murray, A. G. (2009) Using simple models to review the application and implications of different approaches used to simulate transmission of pathogens among aquatic animals. Prev. Vet. Med. 88:167-177.

Raymond, „J.W Rogers, TN Shonnard DR Kline ,AA (2001) "A review of structure-based biodegradation estimation methods" *J. Hazard. Mat.* 84, 189

Schmitt, W (2008). General approach for the calculation of tissue to plasma partition coefficients. *Toxicol. In Vitro* 22: 457-67

Riley RJ, McGinnity DF, and Austin RP (2005) A unified model for predicting human hepatic, metabolic clearance from *in vitro* intrinsic clearance data in hepatocytes and microsomes. *Drug Metab Dispo*s 33: 1304–1311.

Peck, S. L. (2004). Simulation as experiment: a philosophical reassessment for biological modeling. *TREE* 19: 530-534.

Piou, C., Berger, U. and V. Grimm. 2009. Proposing an information criterion for individual-based models developed in a pattern-oriented modeling framework. *Ecological. Modeling*. 220: 1957-1967.

Yu, G.P., Bi, W.Z., Si, D.G., Yang, Y.X., Aisa, H.A., Liang-Zong Xu, L.Z., (2009). Synthesis and QSAR studies on 1-[(5-substituted-1,3,4-thiadiazol-2-yl)methyl]-1H-1,2,4-triazole as antifungal agents. *Struct. Chem*. 20, 569–576.