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

OFFICE OF
CHEMICAL SAFETY AND
POLLUTION PREVENTION

Memorandum

SUBJECT: Transmittal of Meeting Minutes and Final Report for the Federal Insecticide, Fungicide and Rodenticide Act, Scientific Advisory Panel (FIFRA SAP) Virtual Meeting held on September 15-18, 2020

TO: Edward Messina, Esq.
Acting Office Director
Office of Pesticide Programs

FROM: Tamue Gibson, MS
Designated Federal Official
Peer Review and Ethics Branch
Mission Support Division
Office of Program Support

Gibson, Tamue
Digitally signed by Gibson, Tamue
Date: 2020.12.15 12:36:33 -05'00'

THRU: Steven Knott, MS
Branch Chief
Peer Review and Ethics Branch
Mission Support Division
Office of Program Support

STEVEN KNOTT
Digitally signed by STEVEN KNOTT
Date: 2020.12.15 14:39:16 -05'00'

Jacqueline Mosby
Divison Director
Mission Support Division
Office of Program Support

JACQUELINE MOSBY
Digitally signed by JACQUELINE MOSBY
Date: 2020.12.16 08:56:36 -05'00'

Please find attached the meeting minutes and final report for the FIFRA Scientific Advisory Panel public virtual meeting held via phone and webcast on September 15-18, 2020. This report addresses a set of scientific issues being considered by the U.S. Environmental Protection Agency regarding the Office of Pesticide Programs' Use of New Approach Methodologies (NAMs) to Derive Extrapolation Factors and Evaluate Developmental Neurotoxicity for Human Health Risk Assessment.

Attachment

cc:
Alexandra Dunn
Tom Tyler
David Fischer
Carol Ann Siciliano
Cheryl Dunton
Anna Lowit
Dana Vogel
Monique Perron
OPP Docket

FIFRA Scientific Advisory Panel Members

Jeffrey R. Bloomquist, PhD
Robert E. Chapin, PhD (Chair)
Gaylia Jean Harry, PhD
Rebecca L. Smith, DVM, PhD
Clifford P. Weisel, PhD
Raymond S.H. Yang, PhD

FQPA Science Review Board Members

Veronica J. Berrocal, PhD
Harvey J. Clewell, PhD, DABT, FATS
Holly Davies, PhD
Marion F. Ehrich, RPh, PhD, DABT, ATS
Penelope A. Fenner-Crisp, PhD
Helena T. Hogberg, PhD
David A. Jett, PhD
Mark S. Johnson, PhD
John C. Lipscomb PhD, DABT, FATS
Kenneth Portier, PhD
Daniel M. Quinn, PhD
Pamela J. Spencer, PhD, DABT
Lisa M. Sweeney, PhD, DABT, CHMM

FIFRA Scientific Advisory Panel
Meeting Minutes and Final Report
No. 2020-02

Peer Review of the Use of New Approach Methodologies (NAMs) to Derive Extrapolation Factors and Evaluate Developmental Neurotoxicity for Human Health Risk Assessment

September 15-18, 2020

FIFRA Scientific Advisory Panel Meeting

Held via Phone and Webcast
(Virtual Meeting)

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 (FACA) 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 U.S. Environmental Protection Agency (EPA or Agency) Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The SAP serves as a primary scientific peer review mechanism of the EPA, Office of Pesticide Programs (OPP), and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. The 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 and final report are provided as part of the activities of the FIFRA SAP.

The FIFRA SAP carefully considered all information provided and presented by the Agency, as well as information presented by the public. The minutes represent the views and recommendations of the FIFRA SAP and do not necessarily represent the views and policies of the Agency, 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 and final report do not create nor confer legal rights nor impose legally binding requirements on the EPA or any other party. The meeting minutes and final report of the September 15-18, 2020 FIFRA SAP meeting represent the SAP's consideration and review of scientific issues associated with "The Use of New Approach Methodologies (NAMs) to Derive Extrapolation Factors and Evaluate Developmental Neurotoxicity For Human Health Risk Assessment." Steven Knott, MS, FIFRA SAP Executive Secretary, reviewed the minutes and final report. Robert E. Chapin, PhD, FIFRA SAP Chair, and Tamue Gibson, MS, FIFRA SAP Designated Federal Official, certified the minutes and final report, which is publicly available on the SAP website http://www.epa.gov/sap under the heading of "Meetings" and in the public e-docket, Docket No. EPA-HQ-OPP-2020-0263, 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/sap. Interested persons are invited to contact Tamue L. Gibson, MS, SAP Designated Federal Official, via e-mail at gibson.tamue@epa.gov.

TABLE OF CONTENTS

FIFRA Scientific Advisory Panel
Meeting Minutes and Final Report
No. 2020-02

Peer Review of the Use of New Approach Methodologies (NAMs) to Derive Extrapolation Factors and Evaluate Developmental Neurotoxicity for Human Health Risk Assessment

September 15-18, 2020

FIFRA Scientific Advisory Panel Meeting

Held via Phone and Webcast
(Virtual Meeting)

Robert E. Chapin, PhD
FIFRA SAP Chair
FIFRA Scientific Advisory Panel

Robert E Chapin

Date: 14 Dec 2020

Tamue L. Gibson, MS
Designated Federal Official
Office of Program Support

Gibson, Tamue Digitally signed by Gibson, Tamue Date: 2020.12.15 11:57:46 -05'00'

Date:

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
September 15-18, 2020

Peer Review of the Use of New Approach Methodologies (NAMs) to Derive Extrapolation Factors and Evaluate Developmental Neurotoxicity for Human Health Risk Assessment

PARTICIPANTS

FIFRA SAP Chair

Robert E. Chapin, PhD, Independent Consultant, Preston, Connecticut

Designated Federal Official

Tamue L. Gibson, MS
Biologist and Designated Federal Official, Office of Science Coordination and Policy, U.S. Environmental Protection Agency, Washington, District of Columbia, Tel: (202) 564-7642, Email:gibson.tamue@epa.gov

FIFRA Scientific Advisory Panel Members

Jeffrey R. Bloomquist, PhD
Professor, Department of Entomology and Nematology, University of Florida
Gainesville, Florida

Gaylia Jean Harry, PhD
Group Leader, Neurotoxicology Laboratory, National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina

Rebecca L. Smith, DVM, PhD
Assistant Professor, Department of Pathobiology, College of Veterinary Medicine
University of Illinois at Urbana-Champaign, Urbana, Illinois

Clifford P. Weisel, PhD
Professor, Environmental and Occupational Health Sciences Institute, Rutgers University
Piscataway, New Jersey

Raymond S.H. Yang, PhD
Professor (Emeritus), College of Veterinary Medicine and Biomedical Sciences
Colorado State University, Fort Collins, Colorado

FQPA Science Review Board Members

Veronica J. Berrocal, PhD
Associate Professor, Department of Statistics, University of California
Irvine, California

Harvey J. Clewell, PhD, DABT, FATS
Principal Consultant, Ramboll US Corporation, Raleigh, North Carolina

Holly Davies, PhD
Senior Toxicologist, Washington State Department of Health, Tumwater, Washington

Marion F. Ehrich, RPh, PhD, DABT, ATS
Professor, Pharmacology/Toxicology, Virginia Maryland College of Veterinary
Medicine, Virginia Polytechnic Institute and State University, Blacksburg, Virginia

Penelope A. Fenner-Crisp, PhD
Independent Consultant, North Garden, Virginia

Helena T. Hogberg, PhD
Deputy Director, Center for Alternatives to Animal Testing (CAAT), Johns Hopkins
Bloomberg School of Public Health, Baltimore, Maryland

David A. Jett, PhD
Program Director, National Institute of Neurological Disorders and Stroke
Bethesda, Maryland

Mark S. Johnson, PhD
Director of Toxicology, U.S. Army Public Health Center
Aberdeen Proving Ground, Maryland

John C. Lipscomb PhD, DABT, FATS
Senior Toxicologist, CTEH, LLC, North Little Rock, Arkansas

Kenneth Portier, PhD
Independent Consulting Biostatistician, Athens, Georgia

Daniel M. Quinn, PhD
Professor (Emeritus) & Adjunct Professor, Department of Chemistry
University of Iowa, Iowa City, Iowa

Pamela J. Spencer, PhD, DABT
Vice President, Regulatory, Product Stewardship & Quality
ANGUS Chemical Company, Buffalo Grove, Illinois

Lisa M. Sweeney, PhD, DABT, CHMM
Risk Assessment Toxicologist, UES, Inc., Dayton, Ohio

LIST OF ACRONYMS AND ABBREVIATIONS

AC50	Activity Concentration at 50% of Maximal Activity
AChE	Acetylcholinesterase
AChEI	Acetylcholinesterase Inhibition
AED	Administered Equivalent Dose
AIC	Akaike Information Criterion
AUC	Area Under the Curve
BMD	Benchmark Dose
CCA	Comparative Cholinesterase Assays
ChE	Cholinesterase
CNS	Central Nervous System
CV	Coefficient Variation
DAT	Dopamine Transporter
DDEF	Data Derived Extrapolation Factors
DIV	Days In vitro
DMSO	Dimethyl Sulfoxide
DNT	Developmental Neurotoxicity
EFSA	European Food Safety Authority
EPA	United States Environmental Protection Agency
HCI	High Content Imaging
HEC	Human Equivalent Concentration
hNP1	Human Alpha-Defensin-1
HSUS	Humane Society of the U.S.
HTS	High-Throughput Screening
HTTK	High-Throughput Toxicokinetic
NAM	New Approach Methodology
IVIVE	In vivo-In vitro extrapolation
MEA	Microelectrode Arrays
NFA	Network Formation Assay
NOAEL	No-Observable-Adverse-Effect level
NOG	Neurite Outgrowth
OED	Oral Equivalent Dose
OP	Organophosphates
OPP	Office of Pesticide Programs
ORD	Office of Research and Development
PBPK	Physiologically Based Pharmacokinetic
PCRM	Physicians Committee for Responsible Medicine
PD	Pharmacodynamic
PETA	People for the Ethical Treatment of Animals
PK	Pharmacokinetic
POD	Point of Departure
RBC	Red Blood Cells
RfD	Reference Dose
SAS	Statistical Analysis System

SERT	Serotonin Transporter
SD	Standard Deviation
TTC	Threshold of Toxicological Concern
UF	Uncertainty Factor

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel completed its review of the set of scientific issues being considered by the Environmental Protection Agency regarding the Use of New Approach Methodologies (NAMs) to Derive Extrapolation Factors and Evaluate Developmental Neurotoxicity For Human Health Risk Assessment. Advanced notice of the meeting was published in the Federal Register on June 17, 2020. The peer review public virtual meeting was held September 15-18, 2020. The Agency position paper, charge questions, and related documents in support of the SAP meeting are posted in the public e-docket at http://www.regulations.gov (ID: EPA-HQ-OPP-2020-0263). Robert E. Chapin, PhD, chaired the meeting. Tamue L. Gibson, MS, served as the Designated Federal Official.

In preparing these meeting minutes and final report, the Panel carefully considered all information provided and presented by the Agency presenters, as well as information presented by public commenters. The meeting minutes and final report address the information provided and presented at the meeting, especially the Panel response to the Agency charge.

The U.S. EPA presentations were provided during the FIFRA SAP meeting by the following (listed in order of presentation):

September 15-18, 2020: Summary of Meeting Agenda

Day 1 – September 15, 2020

Opening of Meeting – Tamue L. Gibson, MS, Designated Federal Official, EPA, Office of Chemical Safety and Pollution Prevention (OCSPP), Office of Science Coordination and Policy (OSCP)

Introduction and Identification of Panel Members – Robert E. Chapin, PhD, FIFRA SAP Chair

Greetings and Introduction from the Office Director and Division Director– Edward Messina, Acting Office Director, EPA, Office of Pesticide Programs (OPP); Dana Vogel, Division Director, EPA, OPP, Health Effects Division (HED)

OPP Technical Presentation –Introduction and Overview of the Regulatory Use of the New Approach Methodologies (NAMs) – Monique Perron, ScD, EPA, OPP, HED

Welcome – Alexandra Dapolito Dunn, Esq, Assistant Administrator, EPA, OCSPP

OPP Technical Presentation (Continued) –Introduction and Overview of the Regulatory Use of the New Approach Methodologies (NAMs); Developmental Neurotoxicity (DNT) Guideline and Regulatory Context for Organophosphate (OP) Case Study – Monique Perron, ScD, EPA, OPP, HED

Introduction to DNT NAM Assay Development and the US EPA Network Formation Assay – Tim Shafer, PhD, EPA, Office of Research and Development (ORD), Center for Computational Toxicology and Exposure (CCTE)

US EPA High Content Imaging (HCL) Cellular Event Assays for Assessing Chemical Effects on Neurodevelopment Processes – Joshua Harrill, PhD, EPA, ORD, CCTE

Overview of International DNT NAMs Efforts – Tim Shafer, PhD, EPA, ORD, CCTE

DNT-NAMs: Fit-For-Purpose, Results with Organophosphates and Administered Equivalent Dose Comparison to *In Vivo* Benchmark Doses for Acetylcholinesterase Inhibition – Katie Paul Friedman, PhD, EPA, ORD, CCTE

Use of *In Vitro* Acetylcholinesterase Inhibition Data to Develop Data- Derived Extrapolation Factors – Monique Perron, ScD, EPA, OPP, HED

Day 2 – September 16, 2020

Opening of Meeting – Tamue L. Gibson, MS, Designated Federal Official, EPA, OCSPP, OSCP

OP *In vitro* Inhibition Program: Introduction to Testing Program – Richard Reiss, ScD, GVP, Principal Scientist, Exponent

Experimental Procedures and Results – Janice Chambers, PhD, Professor, Mississippi State University

Statistical Analysis of Data – Kelly Higgins, PhD, Senior Scientist, Exponent

Results of Supplemental Variability Study – Richard Reiss, ScD, GVP, Principal Scientist, Exponent and Kristin Lennox, PhD, Managing Scientist, Exponent

Biological Understanding of Interspecies and Intraspecies Variability –Rudy Richardson, ScD, Professor, University of Michigan

PUBLIC COMMENTERS

Oral statements were presented by:

1) Richard Reiss, ScD, On behalf of the OP Coalition of Registrants

2) Kristie Sullivan, MS, Physicians Committee for Responsible Medicine (PCRM)

3) Anna van der Zalm, MS, People for the Ethical Treatment of Animals (PETA)

Written statements were provided by:

1) Vincent Cogliano, Deputy Director for Scientific Programs, California Office of Environmental Health Hazard Assessment and Karen Morrison, Assistant Director and Chief Science Advisor, California Department of Pesticide Regulation

2) Amy Clippinger, Vice President of Regulatory Testing, PETA; and Kristie Sullivan, Vice President for Research Policy, PCRM

3) Vicki Katrinak, Manager, Research and Testing, Animal Research Issues, Humane Society of the U.S. (HSUS); and Gillian Lyons, Senior Regulatory Specialist, Humane Society Legislative Fund

4) Rudy J. Richardson, Exponent on behalf of AMVAC et al.

5) Anne Loccisano and Rick Reiss, Exponent on behalf of Dow Agrosciences LLC

6) Rick Reiss and Benjamin Davis, Exponent on behalf of the Coalition of OP Registrants

7) Kristin Lennox and Rick Reiss, Exponent on behalf of FMC Corporation, AMVAC Chemical Company and Gowan Company

8) Rick Reiss, Exponent on behalf of Coalition of OP Registrants

EXECUTIVE SUMMARY

New Approach Methodologies (NAMs) for Developmental Neurotoxicity (DNT)

Developmental neurotoxicity refers to "any adverse effects on the normal development of the nervous system structure or function". The current EPA DNT guideline study requires an assessment of motor and sensory function, learning and memory, and neuropathology following maternal exposure. The EPA has shifted its testing focus from the developmental neurotoxicity guideline study to more targeted testing due to several challenges associated with the study and its limited impact on human health risk assessments for pesticides. The Agency stated a long-term goal of replacing the DNT guideline due to shortcomings and challenges that have been identified with guideline study data obtained for pesticides. As suggested by the Agency, and inherent in all types of experiments, the quality of these data sets depends on multiple factors that require appropriate expertise and oversight in how the experiments are conducted to control for factors that can increase variance and thus, hinder data interpretation. This has resulted in data submitted to the Agency with a high level of variability. The Panel was not provided information related to the quality and variance of the *in vivo* DNT guideline data sets received by the Agency and only a few of the Panel members had any level of knowledge of the specifics of these assays and the issues of concern. However, the Panel found it clear that the Agency has had issues with DNT guideline study data submitted in meeting their expectations for interpretation of risk.

The Agency considers new approach methodologies (NAMs) that represent non-animal technology, methodology, approach, or combination thereof will provide an opportunity to overcome some of these challenges by evaluating underlying critical processes of neurodevelopment and incorporating human relevant information. These NAMs are considered by the Agency as "something more efficient, less animal intensive, and importantly more human relevant" and "easier to interpret." The Agency made a decision to follow a phenotypic approach for establishing NAMs for developmental neurotoxicity, not unlike the approach taken of apical endpoints for the *in vivo* assessments. The suite of NAMs applied in this manner investigate a number of cellular endpoints likely to be involved in neurotoxic effects, developmental and across the lifespan. In the assay development, important neurodevelopmental processes were modelled with cell-based phenotypic endpoints. Evaluation of specific NAMs for neurotoxicity have been undertaken to determine if they can be used to supplement or to replace existing toxicity tests. The Panel was asked to provide guidance on establishing confidence in the use of the data obtained from the specific NAMs presented. Feedback was solicited with regards to strengths and limitations, adequate reflection of the relevant neurodevelopmental biology, and sufficient development and validation such that data may be incorporated into the assessment of a chemical's effects on neurodevelopment.

The Panel understood the policy requirements of the EPA to move beyond animal testing. The Panel noted that the isolated culture systems lack some features that are known to be critical in the development of the nervous system. The Panel identified numerous limitations and points for consideration that applied to the first three charge questions. It must be acknowledged that given the complex nature of development and the gaps in our current knowledge on these processes, *in vitro* assays may not be representative of many processes and mechanisms that could cause developmental neurotoxic events. These are isolated cell cultures that do not include

inter-organ/tissue communication/effects, peripheral/central influences, and, while powerful in examining specific mechanisms, they may not reflect *in vivo* condition. Statistical differences between exposures in culture may not be representative of actual *in vivo* effects. In the Agency's presentation it was stated that "Decision makers need to understand what endpoint we're measuring, how it is measured, and how that relates to changes and neurodevelopment *in vivo*." In the EPA Issue Paper, it was expected that "Incorporating a battery of NAMs to the evaluation of DNT would also aid in the data interpretation by providing multiple lines of evidence that may help elucidate the biological processes underpinning the apical endpoints affected in the guideline studies." The Panel noted that it was not clear how the NAMs, as stated, will provide such information. Specificity of the target site in the nervous system and much of the circuitry involved in each of the *in vivo* DNT endpoints was not represented in the NAM.

Neurotoxicity does not only occur with direct exposure of the brain to chemicals, it also occurs due to secondary effects from changes in the peripheral environment. For developmental neurotoxicity, these effects can be due to changes that occur in the peri-partum environment (e.g., maternal) as well as due to compromised function of the peripheral organ systems. Specific concerns include:

a. The absence of hormonal factors (sex hormones, thyroid, stress hormones)
b. The influence of neurotransmitter signaling
c. The influence of chemical-induced systemic changes (e.g., inflammation, oxygen levels and distribution)
d. The influence of maternal factors (maternal infection, hormonal, organ system dysfunction, placenta integrity)

In addition, the *in vitro* assays:

a. Will be limited in their ability to detect adaptive or compensatory processes
b. Do not account for critical cell-cell interactions required during neurodevelopment
c. Have difficulty distinguishing between neuroactive and neurotoxic compounds
d. Do not reflect human genetic diversity when using human cell lines from one human

As presented, the NAMs do not include all the different cell types critical during neurodevelopment.

It was not clear to the Panel that a median response in the *in vitro* assays could be directly related to a meaningful point of departure useful in predicting a disease state in humans.

Generally, the Panel thought that these assays constituted an excellent screen, but wondered about their utility in their proposed use to ultimately define a safe level of exposure.

As the assays are performed it was not evident to most Panel members that they would contribute to any understanding of mechanisms.

The Panel considered the efficacy of the *in vivo-in vitro* extrapolation (IVIVE) approach, and the various assumptions involved in its use, for projecting from NAM-derived administered equivalent dose (AED) values to OP doses that inhibit acetylcholinesterase (AChE) in humans and rats. The Panel noted that the development of the NAM technology and the consequent high-throughput toxicokinetic (HTTK) model for IVIVE represented considerable advances

toward the goal of eventually eliminating evaluation of environmental toxicants, specifically organophosphorus (OP) inhibitors of AChE, in laboratory animals. However, the Panel raised a number of concerns. Prominent among the concerns was the adequacy of the HTTK model to account for predictive OP dosing, since extrahepatic metabolic mechanisms are not taken into account. Another concern was the use of data in one species for IVIVE in another. Various recommendations for addressing these and other concerns are detailed in the following discussion.

The analysis does not follow EPA guidance on Data Derived Extrapolation Factors, and the justification for the type (tissue, lifestage, ethnicity) and number of samples analyzed does not support an acceptable level of confidence regarding the breadth of the analysis. These and other conditions complicate the acceptance of presently proposed values for inter- and intraspecies toxicodynamic variability. Data collected to date may be reinterpreted to satisfy some, but not all the issues identified in this review.

The Panel noted that the approach to model-fitting represents current good statistical practice and seems well thought out. Use of an Akaike Information Criterion (AIC) is a typical measure used to choose among different models. Graphical diagnostics are used extensively to supplement AIC in final model selection. The Statistical Analysis System (SAS) code aligns with the description provided in the EPA Issue Paper but includes few comment statements other than indicating which scenario was being addressed and which model was being fitted.

Justifications were not provided as to why default model fitting options were changed for the model being fitted. The logic and/or analysis approach was not presented or documented but has to be inferred by looking at which fit options have been changed from the defaults.

The Exponent analysts expressed proper concern and understanding of the importance of the model fit warnings output by SAS©. The Panel observed that The approach used by the Agency's statistical contractor, ICF to address these model fit concerns, namely the incorporation of parameter scaling, setting the maximum number of search iterations higher, and the use of "ridge" estimation, was reasonable and represents good statistical practice. Most of the model issues were likely tied to a lack of representativeness in available samples and too few sample data points for the model being attempted. The Panel suggested a few additional analysis to address model fit warning issues in the future. For some scenarios the outlier issue reflects the fact that the underlying data display characteristics that are in conflict with the assumptions of the model.

Multiple Panel members agreed that the stratified analyses used to derive values for the bimolecular rate constant ratios employed sample sizes too small to guarantee reliable results. In particular, several Panel members noted the limited availability or absence of samples representing certain ethnic and racial groups, as well as an over-representation of some demographic groups over others, specifically juveniles and infant, compared to the US general population. Several Panel members recognized the supplemental efforts undertaken by ICF to address issues related to model fit. However, Panel members noted that due to the small sample sizes, several issues still persisted, including convergence problems, standard errors estimated to be equal to zero, and extreme outliers.

Several Panel members recommended that efforts be undertaken to increase the sample size, leveraging on planned research efforts at Mississippi State University under the sponsorship of a consortium of companies. A Panel member noted a large variability in the bimolecular rate constant values relative to Naled (a proinsecticide that is known to break down in aqueous media to dichlorvos, a high volatility compound), and indicated possible reasons for the extreme variability in data relative to Naled.

Some members of the Panel believed that from a conceptual point of view, an analysis that uses replicate data to characterize human variability in response to organophosphate exposure was the correct approach. The same Panel members also considered a linear mixed model approach and the IntraClass Correlation the right metric to quantify the extent of human variability in response to organophosphate exposure compared to the total variability (human + experimental variability).

All Panel members recognized that the current replicate analyses use a very small sample size, resulting in tests that are underpowered. Due to the very small sample size, a Panel member recommended that coverage ratios for the bimolecular rate constants k_i be calculated using a different approach than that used currently, which relies on approximations due to large sample size.

The Panel was supportive of the idea of additional replicated data analysis to characterize the sources of variability in the human response to exposure to different organophosphates. The Panel recommended that a larger sample size with samples that are more representative of the US general population be used in these additional replicated data analyses. Some Panel members recommended a better accounting for all sources of uncertainty, such as handling the differences between chemical properties of organophosphates, the nature of the data (not actual observed data, but derived quantities, point estimates), and the uncertainty associated with extrapolation of cellular responses to responses of whole biological systems.

PANEL DISCUSSION AND RECOMMENDATIONS – New Approach Methodologies for Developmental Neurotoxicity

Charge Question 1. – ***New Approach Methodologies for Developmental Neurotoxicity***

Question 1. *For charge questions 1-3, the overall focus is on the ability of the developmental neurotoxicity (DNT)-new approach methodologies (NAMs) to evaluate important biological processes related to neurodevelopment. EPA is soliciting feedback on whether the NAMs adequately reflect the biology such that data may be incorporated into the assessment of a chemical's effects on neurodevelopment.*

Using primary rat cortical neurons grown on microelectrode arrays (or MEAs), the EPA's Office of Research and Development has developed a network formation assay (NFA) to assess the potential impact of chemical exposure on neural network formation and function as described in Sections 2.3.1 - 2.3.4 of the Agency's Issue Paper. *Please comment on the strengths and limitations of using this assay to evaluate the biology underlying network formation as a component of neurodevelopment that may be susceptible to modulation by chemical exposure.*

The neuronal system for network formation assay (NFA) presented by the EPA was rodent cortical neurons. These cells form densely organized cellular networks and become electrically active after a few weeks in culture. Typically, differentiated neurons are comprised of a cell body from which neurites (dendritic tree) and a single long cylindrical axon emerge. These structures adhere to MEAs substrates by electrostatic or chemical interactions between adhesion molecules that protrude from the lipid membrane of the neurons and molecules deposited on the MEA platforms. An ionic solution fills the cleft between the cell membrane and the MEA substrate. In Section 2.3.1. of the EPA Issue Paper, the assay defines relevant signaling endpoints as parameters that indicate general and bursting activity as well as a calculated network spikes. The MEA methodology has been around for a number of years and, with technological advancements as well as data capture and analysis, the commercial systems offer an attractive method to try to examine neuronal connectivity *in vitro* to study cell-cell communication.

Strengths:

Rapid, high-throughput screening (HTS) format; functional endpoint of neuronal activity; ability for repeated measures over time; ability to compare acute responses versus disruption of cell development with exposure.

Several Panel members noted strengths of the NFA including assessment of functional endpoints that could serve as descriptive apical endpoints, the ability to examine different stages of exposure, and the ability to record repeated measures over time.

The NFA allows for a level of limited high-throughput, making it suitable for screening compounds. One Panel member noted that the MEA NFA often finds effects at or below the lower quartile of the Toxcast/Tox21 activity concentration at 50% of maximal activity (AC50) values (US EPA 2020b, pages 8-9).

The Panel noted additional strengths of the MEA NFA in that the anatomical features of the cells can be further examined by incorporating additional imaging and/or immunohistochemical methods. Comparison of effects that are dependent upon prolonged exposure, DIV age dependent exposure, acute chemical presence, could provide information on developmental versus neurotoxic/active effects. The ability to examine cell viability, growth, maturation, in addition to spontaneous activity can provide information for data interpretation. The NFA has the ability to expand to include targeted neuron to neuron signaling and allows for supportive information on alterations in neuronal network signaling.

The NFA assay as described was amenable to further examination to integrate with other endpoint assessments such as molecular profiling or biochemical endpoints with pharmacological modulations to identify the "underlying biological effects". With optimization, the MEA NFA paradigm can be used to examine different neuronal populations.
The application of rodent cortical cells has the benefit of a deep literature base for the *in vitro* establishment and use of these cells in neurobiology. Additionally, they represented a primary culture containing neurons and astrocytes which avoids constraints associated with cell lines or with manipulations required to differentiate cells into neurons. The use of rodent cells allows for the generation of cells from males or females and for the design of future studies to evaluate and predict validity of effects observed *in vivo*. No data was presented to compare sensitivity across species to support usage of any specific cell source. The inclusion of astrocytes was a strength. However, the assay design does not allow for distinguishing if effects on network formation are related to effects on neurons or astrocytes.

In vivo, circuitry development and orientation relies heavily on the lamination of brain structures (e.g., cortex, hippocampus, cerebellum) and cell-cell signaling. In contrast, *in vitro* neural network formation depends on cell specific recognition cues rather than orderly lamination to promote specificity of synaptic connectivity (Williams et al., 2011). The random nature of the network may not be a critical factor for screening; however, the Panel recommended caution for translating to brain development.

Limitations:

Difficulty detecting adaptive changes; lack of inclusion of neuroendocrine/neurotransmitter interactions; difficulty distinguishing neuroactive from neurotoxic compounds; high variability. Several Panel members noted the MEA NFA, such as the high content imaging (HCI) assays, lacked any contribution of hormone levels and/or changes in neurotransmitters. This gap can impact the ability of the assay to detect adaptive or compensatory processes. Moreover, the assay will have difficulty distinguishing between neuroactive and neurotoxic compounds (i.e. lead to more false positives).

Assay variability was raised by several Panel members who considered the coefficient variation (CV) of the assay parameters as too high (often >15%). Another Panel member thought it reflected appropriate variation based on the biological complexity of the measurement. The Panel noted that more targeted data will be required to determine if this level of variability is inherent in the biological endpoint being measured.

The Panel considered that, as a standalone test, the MEA NFA will not detect all potential DNT compounds as the target (e.g. early windows of development or certain cell types (e.g. dopaminergic neurons and oligodendrocytes)) might not be present. However, the Panel generally agreed that the assay, if used in conjunction with a battery of assays (i.e. DNT NAMs), is an appropriate screening tool for neuronal function.

Other Limitations:

One Panel member commented that many compounds will show a neuro-excitation effect at low concentrations and fast nerve block at high concentrations. Due to this, the Panel considered the question of which measured effect was most appropriate for evaluating assay performance. The Panel considered the need for additional known positive controls with domoic acid. It was noted by the same Panel member that, by including more specific neuroactive compounds, additional positive effects on general activity, bursting activity etc. would have addressed a current deficiency (US EPA 2020a, page 30).

Several Panel members cautioned that an *in vitro* test system does not sufficiently represent the *in vivo* environment nor do changes *in vitro* translate to effects on development. The NFA, measured signaling sites and spontaneous activity however, translating this to representing mature synaptic signaling across defined neuronal networks was recommended to be done with a high level of caution. A critical process of network formation is related to the stages where neurons undergo a phase of overconnectivity, followed by synaptic pruning and thus, a refinement of network formation and activity. The exposure paradigm of the NFA does not take this developmental process into consideration.

One Panel member commented that there was a need for empirical data that showed *in vitro* concentrations represented relevant *in vivo* nervous system concentrations to cause developmental neurotoxic effects. In some cases, it has been compared with blood and brain acetylcholinesterase (AChE) activities, but not developmental toxicity. *In vitro* to *in vivo* extrapolation (IVIVE) was considered as a tool to help inform if an *in vitro* change was sufficiently representative of a plasma concentration that could lead to an adverse outcome.

Panel members offered the following comments related to concerns with underlying assumptions or enhancing the experimental paradigm of the assay:

- Trying to control or standardize the cultures for cell density with prolonged exposure may be difficult if processes in the earlier days *in vitro* (DIV) were affected. This could affect data interpretation in that the resulting differences in network signaling could represent a difference in cell maturation or a difference in cell signaling ability. While standardization approaches were mentioned by the Agency, unless this is a one-to-one relationship, the standardization techniques might not reflect the biological differences.
- Discriminating between effects on cell signaling occurring as a result of exposure over DIV, or effects occurring within the earlier DIV, versus effects occurring due to the presence of the chemical was not possible given the experimental design presented. However, the assay exposure paradigm could be modified. For example, excluding the early time point, cells at each of the DIV could be examined for the "acute" effect of the chemical to determine if any differences attributed to network signaling are due to early DIV exposure or to the presence of the chemical at the time of assay.

- Synchronized activity of rodent cortical neurons initiates at approximately embryonic day 16, increasing in frequency, then subsiding by first week of life (Corlew et al., 2004). It is thought that the acquisition of synchronous firing is a key property in the development of cortical neural networks. Further inclusion of synchronous firing might allow for examination of the pattern of network development over DIV.
- The assay paradigm makes the assumption that DIV of newly obtained neurons from the postpartum rodent brain recapitulates the brain development process.
- In analysis of the data, the interdependency of many of these endpoints and the complexity of the system may require a different framework than what has been done in the ToxCast arena. Since 2018, there have been a number of papers published dealing with the use of a deep learning framework for classifying data from MEA. Many of these approaches are designed to address the biology underlying the MEA data and to consider how to examine the data in a manner that will decrease variability (Buccino et al., 2017; 2018a,b).

Beneficial asymptomatic seizures are important in brain development, especially during pruning. Determining how the MEA NFA can be designed to detect chemicals with seizurogenic properties or those that may inhibit seizure activity will be of importance in data interpretation. Several Panel members commented on the need to ensure the quality and reproducibility of the data between laboratories through well-developed standard operating procedures (SOPs) for designing and conducting the studies. In the case of the EPA's Office of Research and Development (ORD), they are fortunate to have the level of expertise to carry out robust standard operation procedures (SOP) formulation.

Recommendations:

The Panel recommended a number of avenues for developing the use of the MEA NFA to provide information into the realm of "developmental neurobiology/toxicology" that would be of benefit to the Agency and use of the data for risk assessment. These include:
1) Opportunities to expand and develop the MEA NFA further, with a focused effort, given the potential to be able to integrate this assay with other endpoint assessments (e.g. molecular profiling to try to identify "underlying biological effects").
2) Deeper, targeted examinations of the associated underlying mechanism using either a more complex approach or proposed deep learning analysis of the data.
3) Modifications of the assay to include additional aspects of response rather than relying solely on spontaneous activity.
4) Electrochemical impedance spectroscopy to study cell adhesion and growth,
5) Pharmacological methods to modify the chemical response
6) Evaluating the specific chemicals that inhibit AchE (ranged from 0.03 micromolar to 10 micromolar), concentration should be considered that are based on AchE inhibition including doses below those that inhibit AchE doses and doses above steady state inhibition.
7) Inclusion of an explanatory figure describing general activity/burst rate (US EPA 2020a, Table 1).

The Panel concluded that the MEA NFA has the potential to go beyond the realm of "screening for chemical prioritization" to developing a more "fit for purpose" approach to identify "underlying biology" of neurodevelopmental processes as relevant to the specific question at

hand. Such an integrated, composite approach could be used to evaluate not only the cell signaling readout for other chemicals but also classification of critical biological events.

Charge Question 2. – ***New Approach Methodologies for Developmental Neurotoxicity***

Question 2. The EPA's Office of Research and Development has used high content imaging (or HCI) with a variety of rat- and human-derived *in vitro* models to investigate the potential impact of chemical exposure on cell proliferation, apoptosis, neurite outgrowth, and synaptogenesis as described in Sections 2.3.1 - 2.3.4 of the Agency's Issue Paper. *Please comment on the strength and limitations of using the HCI assays to evaluate the biological processes underlying proliferation, apoptosis, neurite outgrowth and synaptogenesis as components of neurodevelopment that may be susceptible to modulation by chemical exposure.*

The HCI of neuronal progenitor cell lines represented an assay utilizing human and rat lines to investigate different morphological stages and processes associated with development including cell proliferation, cell death/viability, differentiation, and process outgrowth that reflect aspects of brain development, and phenotypically may be representative of developmental neurotoxicity. The assay relies on new technology for video imaging and quantitation of distinct morphological features of cultured cells.

Strengths:

Covers several endpoints, high-throughput, automated, reproducible. The strengths of this assay are that the cell morphology aspect compliments the cell function aspects of the microelectrode array assays. It was also a method used in documenting ontogeny studies at the cellular level related to DNT from early (proliferation) to late (synaptogenesis). The HCI assays span early and late (embryonic and fetal) neurodevelopmental processes. It is a relatively high-throughput assay where several chemicals can be tested in parallel. The inclusion of human and rodent cells can be useful to identify species differences. The Panel generally agreed that no single *in vitro* screening assay can recapitulate the critical processes of neurodevelopment or affirmatively identify all chemicals that may produce DNT.

The HCI endpoints offer good reproducibility and a more statistically robust evaluation of neurodevelopment than the MEA method, particularly evident in the neurite outgrowth assay in both the rat cortical and hN2 cell lines. One might expect that the two assays would have similar responses to the vehicle dimethyl sulfoxide (DMSO), but that was not the case. There were smaller CV values calculated for the HCI parameters, compared to the MEA with the vehicle DMSO. Many (two thirds) of the HCI CV values were <10%, and only one third were >10%, and no CV values for HCI parameters were >20% (US EPA 2020a, Table 5, pages 26-27). The variability for 21 HCI CV values (mean + standard deviation (SD) = 8.74 $\pm$ 4.5) was significantly less (t-test, p = 0.0002) than the MEA CV values for 19 electrical parameters (CV mean + SD = 16.54 $\pm$ 7.3).

The Panel agreed that one must consider that observed differences can also be due to the cell culture differences such as cell composition, window of development, rate of development, and

media composition (e.g., binding vs. free chemicals might differ). One Panel member commented that the comparison of heatmap clustering (US EPA 2020a, Table 11) appeared to be a useful way to ascertain the weight of evidence (or probability) of a substance to cause developmental neurotoxicity. The suite of endpoints in the HCI assays may be useful as toxicity screening tools, but issues remain regarding their use in protection of developmental neurotoxic effects in humans.

Limitations:

Bias in HCI measurements; discrete assays that do not include other inter-organ interactions (e.g. neuroendocrine); absence of functional endpoint; uncertainty regarding use of AC50 concentrations; limited utility for predicting DNT mechanisms, unclear for broad utility across compounds. Several Panel members were uncertain about the utility of AC50 concentrations and questioned whether they were predictive or translate to DNT effects that occur with concentrations in human plasma associated with an adverse outcome. One Panel member commented that, while instrumentation data capture and analysis software can provide quantitative measures, human evaluation and potential bias is still possible and there are required quality control efforts to confirm that the automated data collection is capturing the specific cell morphology of concern. These are isolated systems where biological changes between systems are not accounted for (e.g., neuroendocrine effects), and the ability to extrapolate findings to in vivo neurodevelopment was a concern for several Panel members.

Many of the HCI endpoints captured phenotypic changes that may occur with multiple underlying modes of action that are likely to affect neurodevelopment. However, they do not provide information on specific mechanisms that may cause neurodevelopmental effects. Some endpoints (or modes of action) that could plausibly cause neurodevelopmental effects are absent (neurotransmitter-specific biomarkers of effect).

Generally, the Panel thought that these assays constitute an appropriate screening tool, but questioned their utility in their proposed use to ultimately define a safe level of exposure.

Other Limitations:

Several Panel members questioned the time of exposure and whether 24-48 hours (5 days for synaptogenesis) of exposure is enough to mimic developmental neurotoxicity *in vivo* beyond acute poisoning. While a window of exposure may exist for damage, the outcome is often associated with either developmental time or long-term exposure.

Another Panel member commented that the synapse formation assay only includes the presynaptic marker synapsin I and recommended that additional immunological markers be employed to determine if a synapse is formed, preferably localization with a postsynaptic marker such as postsynaptic density protein (PSD) 95.

One Panel member mentioned that when the assay-positive control chemicals were selected (US EPA 2020a, Figure 4) it would have been beneficial to show all chemicals over the assays, especially since some of them are evaluating the same endpoint (e.g. neurite outgrowth). This would benefit the understanding of differences and similarities in the various assays and how

they could complement each other or how they provided conflicting data. This approach would also be of value to evaluate the 120 chemicals in the European Food Safety Authority (EFSA) study.

The EPA Issue Paper, page 16, states that "chemicals might contribute to e.g. autism spectrum disorder, attention deficit hyperactivity disorder, though the specific neurodevelopmental mechanisms for these DNT outcomes is the subject of ongoing research efforts." It was not evident to most Panel members that HCI assays would contribute to any understanding of mechanisms of these disorders, especially as many of these outcomes are likely due to gene-environmental interactions. One Panel member commented that sex differences would not be addressed with the assay as described.

Several Panel members commented about the possible biases in HCI interpretation. The ability to automatically perform and analyze large numbers of complex endpoints and to replace human bias in image interpretation by numerical representation of cellular behavior in human and rat cell lines will be challenging. This was evident in Table 6 of the EPA Issue Paper where 2/3 of the HCI endpoints have CVs below 10%. This contrasts with the MEA NFA where nearly 70% of the endpoint CVs are above 10%. Overall robust positive controls for all the activity types measured apart from activity for human alpha-defensin-1 (hNP1) proliferation which produced effects, but of a low magnitude (US EPA 2020a, Figure 4, Table 9). It was recommended by one Panel member that assessing multiple cellular markers in a single assay may be optimal.

Several Panel members noted the lack of inter-organ effects. That they lack endocrine signaling/communication that are known to influence neurodevelopmental processes could lead to false negatives for some classes of chemicals. The assays do not eliminate animal use as some rely on animal derived cell lines (e.g. neurite outgrowth (NOG) rat cortical), though it was recognized that overall animal use will likely be decreased. Both NOG rat cortical and hN2 cell lines appear to be equally sensitive in detecting neurite outgrowth (US EPA 2020a, Table 9) although the positive control, lithium chloride, used for hN2 cell line appear to have some cytotoxicity at the effective concentration. One Panel member questioned the need for both rat cortical NOG assay relative to the hN2 cell line assay and if they showed significant species specific differences to warrant use of both. A direct comparison between the rodent and human cell lines would be a contribution to any future assay design or selection.

One Panel member noted the absence of neurotransmitter-specific biomarkers analyzed by HCI and raised the issue that this absence might cause the HCI analysis to miss this specific type of toxic effect. Such neurochemical markers could include tyrosine hydroxylase, dopamine transporter (DAT), serotonin transporter (SERT). Development and/or synaptogenesis could be skewed by a compound in favor of or against a particular transmitter type without changing overall synaptic density. The HCI assay suffered from a lack of endocrine input on development. The Panel recommended that, by using what is known about cortical neuronal development, there may be specific patterns of ion channel, receptors, or other proteins that could be leveraged as additional key or sentinel measurements of chemical insult, in addition to the general markers of synaptogenesis already proposed.

One Panel member raised the question of whether the HCI assay could be expanded to capture information on axonal and dendritic growth separately given that differential effects have been reported with chlorpyrifos (Howard et al., 2005).

Several Panel members raised questions about the utility of using postnatal cells in an assay that assesses developmental neurotoxicity where effects could occur prenatally however, another Panel member noted that cross species timing for brain development allows for a closer match across rodent and human development for cell harvest. Another Panel member recommended a requirement that the development timing of collection from rodents was determined to be optimal for predictions of *in vivo* neurodevelopmental effects and possible translation to humans. One Panel member raised concerns with making the assumption that days *in vitro* represented the developmental progression *in vivo*.

Several Panel members questioned the relevance of using a median concentration in a statistical test between treatments as the point of departure to be used in an IVIVE extrapolation effort. Since many of the positive controls can be used therapeutically, concentrations in an assay that could result in a disease state (i.e., developmental neurotoxicity) has not been determined. More justification is needed to show what magnitude of *in vitro* response in these assays translates to *in vivo* disorder. Some of the changes can be used therapeutically and may have little relevance in causing DNT at those levels and as such violates the first assumption. However, it was remarked that some of these endpoints could be refined if compared with *in vivo* responses of known effect level.

Recommendations:

The Panel recommended that the Agency consider the following:

1) Linking of morphological endpoints *in vivo* with endpoints of effect at concentrations used in the *in vitro* assays.

2) Determine the concordance of HCI studies between human neural progenitor cells and primary rat cortical neurons to obtain the value of using one versus the other cell source.

3) Inclusion of a vehicle control in every assay and consider the inclusion of a different vehicle substance rather than dimethyl sulfoxide

---

Charge Question 3. – ***New Approach Methodologies for Developmental Neurotoxicity***

Question 3. As discussed in Section 2.1 of the Agency's Issue Paper, EPA has shifted its testing focus from the developmental neurotoxicity guideline study to more targeted testing due to several challenges associated with the study and its limited impact on human health risk assessments for pesticides. New approach methodologies (or NAMs) provide an opportunity to overcome some of these challenges by evaluating underlying critical processes of neurodevelopment and incorporating human relevant information. NAMs covering critical processes in neurodevelopment developed by EPA's Office of Research and Development and

researchers funded by the European Food Safety Authority are presented in Table 3 and Figure 2 of the Agency's Issue Paper (Section 2.3.2). Based on this information and considering the goal of developing a NAM testing strategy or an integrated approach to testing and assessment (or IATA) within the next year for evaluating developmental neurotoxicity to inform chemical risk assessments, *please comment on whether this NAM battery reasonably evaluates the biology underlying the critical processes related to neurodevelopment that may be susceptible to modulation by chemical exposure.*

The Panel complimented the EPA for advancing the development and critical evaluation of NAMs for DNT and for consideration in the Agency's scientific and regulatory processes. The Panel appreciated EPA's efforts to present information on the status of assays being developed in international efforts. A consideration and appreciation that one assay system will not serve to reflect the various aspects of neurodevelopment and impacts of chemical exposure was reflected in the battery of assays presented. The battery represents the EPA assays and assays developed in two European laboratories under funding by the European Food Safety Authority. Other contributions by the Danish EPA were mentioned but not presented in detail. The battery of assays was presented in Section 2.3.2, Figure 2 and Table 3 in the EPA Issue Paper. The assays presented represented three from Dr. Leist's laboratory (Konstanz University, Konstanz Germany) including a neural crest migration assay in which migration was identified by cells moving into an open plate region upon removal of a barrier. The second utilized a human H9 cell line manipulated for differentiation into dorsal root ganglion precursor cells that are then examined for alterations in neurite extension. The third assay was an immortalized human dopaminergic neuronal cell model in which neurite outgrowth was measured. Dr. Fritsche's laboratory (Leibniz Institute for Environmental Medical Research, Dusseldorf, Germany) focused on obtaining multiple endpoint assessments using rather complex neurosphere assays derived from primary human neuroprogenitor cells of fetal origin. This system allowed for evaluation of cell proliferation followed by neuronal and glial (radial and oligodendroglia) differentiation and migration of neurons preceded by radial glia similar to what can be observed *in vivo*. Using various immunological markers, the morphology of cells and various quantitative measures of cell density and migration areas can be determined.

The Agency briefly mentioned that efforts were underway in 5 different laboratories examining the utility of developmental assays in zebrafish. These studies were still ongoing, and data was not available for the Panel to provide an assessment. However, the consideration of the zebrafish model to determine if it can be demonstrated to be specific for effects on nervous system development was supported by the Panel. However, while the zebrafish may offer a method to include various systemic factors, there remain limitations in the translation of hormonal, metabolism, and simplistic behaviors.

Strengths:

Overall, the Panel agreed that the focused battery of assays was an excellent attempt to reflect, if not directly model, critical morphological processes identified as being involved in nervous system development. The individual components of the battery represented assays that could be selected as a fit for purpose and would allow additional endpoint evaluation to allow the Agency to gain a better understanding of the biological processes being recruited. The assays were currently being conducted in laboratories with the level of expertise and instrumentation

required. In addition, there was some level of redundancy of the processes measured that may allow for replication of effects. The International effort to standardize and validate the methods will be a benefit for consideration of use in regulatory decisions.

Limitations:

The Panel pointed out that several important processes and cell types were missing in the battery and that it underestimated the complexity of nervous system development. Nervous system development occurs in a very interactive manner with multiple cell types driving the temporal and spatial progression that leads to a correct formation of the neural network. These interactive processes are difficult to capture in cell models. The assay systems developed in the Fritsche laboratory offer the greatest possibility and the MEA system proposed by the Shafer laboratory offers some level of functional as well as structural assessment. The battery provides primarily structural assessments and overall lacks functional and mechanistic assessments. The phenotypic changes likely reflect functional effects occurring in the cell.

The Panel identified a significant limitation in assays covering glial cells (astrocytes, oligodendrocytes, and microglia). This was considered a major limitation as these cells comprise approximately 50% of the neural cell population and play crucial roles during development. Thus, the NAMs proposed were focused on neurons and exclude the critical functions of various glial populations and the neurovasculature unit in nervous system development. While it is critical that functional neuronal networks are formed, the ontogeny of microglia and astrocytes and their essential contributions to this process was not fully considered in the NAMs presented. The Panel noted the limited representation of the various neuronal populations (neurotransmitter type, and brain regional) and the absence of information on processes that are known to be critical for brain development such as ontogeny of neurotransmitter function (levels, receptor expressions, activation, or neurotransmitter ratio). Some Panel members noted the limitation of the assays to cover differentiation to specific neuronal types, e.g. dopaminergic, cholinergic, and serotonergic neurons and the various nervous systems (central, peripheral, autonomic). Multiple Panel members noted that neurotransmitters are not only important for nervous system function but for several developmental processes cited as critical in the design of current assays such as migration, synapse formation and neurite outgrowth. The Agency's presentation mentioned the possibility of chemicals contributing to autism spectrum disorder or attention deficit hyperactivity disorder for which published research suggested an imbalance between excitatory and inhibitory neurotransmitter signals requiring consideration of a more complex cellular interaction. Thus, for a developmental human disease with possible association with chemical exposure, the assays proposed lack a critical component for assessment.

One Panel member pointed out that the natural *in vivo* progression of neurons sending out processes was not random but rather a targeted migration to a final "correct" location to form a synapse. This migration was driven by chemotaxic and chemoattractant signals. The migration assays included in the battery do not include a directionality or signaling component but rather examine random cell movement out of a dense core for the assays in the Leist laboratory, or process outgrowth in the EPA assays or the neurosphere assay. Based upon data available in the literature, methods to evaluate response to migratory directional signals are available that could be included in the assays. One Panel member commented on the absence of assays that examine the neurovascular unit, either individual cells or as a unit that significantly contributes not only to

almost all aspects of brain development and maturation but also to establishment of the blood brain barrier which is formed during gestation. Another Panel member mentioned the importance of metabolic differences during development and wondered if this could be captured by the battery.

The translation of the various assay endpoints to the *in vivo* manifestation of developmental neurotoxicity remains an issue of concern. Two Panel members strongly supported further research demonstrating a level of corroboration between *in vitro* assay effects and *in vivo* neurotoxicity to better understand the translational relevance from *in vitro* assays to *in vivo* manifestations of neurotoxicity and eventually to human health effects. Both aforementioned Panel members considered that critical questions remain on the applicability of these proposed *in vitro* assays as being representative of critical processes involved in nervous system development and on how well they might replace the *in vivo* rodent studies.

Recommendations:

Several Panel members identified challenges in providing feedback and questioned the readiness of the NAM battery for testing and assessment within the next year as the data for many endpoints will not be available until 2021 and not until 2023 for zebrafish behavioral studies. **Therefore, several Panel members recommended to postpone a final recommendation to a future scientific advisory panel once the data are available as this will weigh heavily on choices made.** Once the data are generated, the Panel acknowledged the importance to compare the results of the whole integrated testing battery with results from *in vivo* human and animal data. This is not meant to preclude the ability of the Agency to utilize all valid and relevant data in their efforts to determine risks for human health.

The Panel considered that any one NAM was insufficient for determining the potential for a chemical to cause neurotoxicity or developmental neurotoxicity. Rather, using a battery of tests with possible redundancies would be a better approach. The Agency asked the Panel to consider how the available data from NAMs may be used in current regulatory activities. In general, the Panel agreed that if the Agency uses published data in their evaluation, then there is no reason to exclude peer-reviewed published *in vitro* assay data - whether screening or mechanistic - in that final "weight of evidence." This could include neurotoxicity assays as well as other relevant published data as it corresponds to specific related biological processes. One point of consideration, however, would be to ensure that a sufficient sample size was included in any such publication.

All of the Panel members recommended that any battery be a “living and evolving process” that can be revised and improved with new technology, assays, information on validity and reliability and *in vivo* translation. One Panel member expressed concern that once a battery of assays is accepted for use in the regulatory arena, that there will be little “living process” and no incentive for developing new assays, refining endpoints, or establishing underlying mechanisms and biological processes involved. It was considered important to develop a robust base with critical evaluation and challenge including reproducibility and transferability between laboratories to avoid the mistakes and later-identified limitations of the DNT *in vivo* guideline studies. One Panel member recommended adopting the Bayesian Approach for this purpose and have it placed prominently in the “EPA Issue paper.” A reference was shared with additional

information on the step-by-step Bayesian calculation and numerous examples (McGrayne 2011). The Panel **recommended that the Agency take a leading role in ensuring this evolving, refining, living process to maximize the utility of the data obtained for regulatory decisions.**

This is not meant to preclude the ability of the Agency to utilize all valid and relevant data in their efforts to determine risks for human health.

Many Panel members recommended that any evaluation of the NAM data include transcriptional profiling, neurotransmitter profiling and measurements of other molecular markers as complements to better understand the biological plausibility of observed effects from the other assays.

Near-Term Actionable Items:

The various laboratories contributing to the NAMs for DNT have generated a large data set for evaluation. While there are efforts to review the data with expected results in 2021, there are specific types of information that will be relevant for any use of data by the EPA for weight of evidence. There was a notable absence of using the data available to conduct a power analysis to determine the number of observations required for confidence in any findings. It was confirmed by the Agency that such an analysis on each of the assay endpoints had not been conducted. The Panel recommended that EPA consider using their available data on control samples for a Power Analysis within any specific culture preparation and across culture preparations.
The Panel considered that further refinement in investigating statistical differences between treatments in these NAMs with *in vivo* endpoints was needed.

The Panel also recommended comparing data from existing *in vitro* screening efforts on uniform cellular responses (e.g., viability, proliferation, maturation) across multiple cell-based *in vitro* systems to develop criteria for identifying a chemical as neurotoxic versus cytotoxic.

Additional Panel Recommendations:

Further refinement in investigating statistical differences between treatments in these NAMs with *in vivo* endpoints is recommended.

Several Panel members were encouraged by the progress made in refining these assays and some thought that they would be useful for screening approaches. However, more work is needed to show that changes in isolated cultures are representative of *in vivo* effects. Efforts to evaluate the *in vitro* assays for their predictive validity of a "neurotoxic" effect in the developing brain are highly recommended and are considered by many Panel members as critical for the interpretation of the data for regulatory decisions. How representative are the endpoints of *in vivo* biological processes? The Panel noted the answer is "the *in vitro* assays are reasonably representative", but the distance between *in vitro* models and *in vivo* nervous system remain large, and are a cause for concern and a focus of future effort.

Consider linking morphological endpoints *in vivo* with endpoints of effect at concentrations used in the *in vitro* assays. In risk assessment applications of toxicity data, considerable attention is applied in understanding the magnitude of uncertainty associated with inter- and intraspecies

variability. Two Panel members commented that there was uncertainty each time they used data collected at one level of biological organization to apply to another (i.e., going from cells to tissue, tissue to individual, and individual to populations). Since NAMs are missing tissue-to-tissue level interactions, it makes sense to also consider an Uncertainty Factor to go from cell/tissue-based responses to individual-level responses. These independent assay results could also be evaluated holistically. Consider an assay sensitivity approach to develop a value that would be protective of various endpoints as measured by various assay endpoints.

Establish a common agreement of what constitutes a neurotoxicological effect in the assays and if the thresholds of concern are considered clinically relevant.

Establish an approach that will be accepted to determine if the tested chemical concentrations reflect target tissue levels and how to integrate that into the evaluation.

In the Agency's presentation it was stated that "Decision makers need to understand what endpoint we're measuring, how it is measured, and how that relates to changes and neurodevelopment *in vivo*." The Panel agreed with the critical need for the decision makers to understand how the NAM endpoint relates to *in vivo* neurodevelopment. To this end the Panel recommended that the Agency actively participate in a collaborative effort to bring interested parties to the table for critical discussion on these assays and the proposed use and regulatory expectations. This would include those who have been actively developing the assays, those who will need to use the data, but also those with *in vivo* knowledge of developmental neurobiology, neuroanatomy, and neuropathology. If it is the goal of the Agency to be able to interpret the data within the framework of a neurotoxic effect and translate that to humans, a broader level of expertise and input is required.

Encourage the development and inclusion of glial based (astrocyte, oligodendrocyte and microglia) and neurovascular based targeted relevant NAMs.

In the EPA's Issue Paper on NAMs for DNT, one of the concerns and considered limitations of the *in vivo* DNT guideline study was "challenges associated with the study and its limited impact on human health risk assessments for pesticides". There are a number of concerns with this request for a direct association to reflect an ever-shifting over-expectation of the *in vivo* data set that are beyond what the basic biology would support and what the original drafters of the guidelines intended. Thus, to minimize the chances of this happening with the shift to NAMs, it is recommended that the Agency lay out the expectations of those making regulatory decisions and to determine if those expectations are well beyond the performance and information obtained from the assays.

Charge Question 4. – ***New Approach Methodologies for Developmental Neurotoxicity***

Question 4. Organophosphate pesticides share the ability to inhibit the acetylcholinesterase enzyme, which prevents the breakdown of acetylcholine leading to neurotoxicity. Inhibition of acetylcholinesterase is the basis of current OP human health risk assessments. In order to compare the relative sensitivity of the MEA NFA and HCI assay results to doses that inhibit acetylcholinesterase in laboratory animals, *in vitro* to *in vivo* extrapolation (or IVIVE) approaches were used to approximate NAM administered equivalent doses for a subset of

organophosphate pesticides as described in Section 2.3.6. *Please comment on the strengths and limitations of this comparison and whether there are alternative approaches for this evaluation using the available data.*

Strengths:

A Panel member noted the importance of toxicokinetic models that can bridge the gap between *in vitro* activity levels and external exposure levels and thus are key to providing EPA Office of Pesticide Programs with risk-based contexts for their policy decisions. The Panel member noted the following strengths: 1) the graphical presentations of the rat and human *in vivo to in vitro* extrapolation (IVIVE) comparisons were very useful visualizations of a lot of information; 2) the use of higher-quality models where available, with comparisons of HTTK and physiologically based pharmacokinetic (PBPK) models, was also a strength; 3) the modeling is generally well documented, with both the general and "simplifying" assumptions explicitly stated up front. The Panelist also outlined several weaknesses: 1) EPA did not address assumptions articulated for the IVIVE approach and HTTK modeling with respect to the organophosphates (OPs) case study; this weakness was identified by at least three Panel members; 2) The performance of the HTTK 3-compartment model was not articulated with respect to "fold-error" for the spectrum of test chemicals or specifically for OPs; model performance standards for differing purposes were not articulated; 3) One alternative approach was used for parameterization of hepatic intrinsic clearance (Cl_{int}); others exist; 4) The limited number of PBPK model comparisons could potentially be expanded.

One Panel member discussed recommendations for addressing weaknesses identified by the Panel. The same Panel member recommended re-examination of assumptions pertinent to assays performed and test chemical, as outlined in the following passages: The EPA appeared to be omitting the groundwork required to build confidence that the IVIVE strategy and tools chosen are applicable to the assay points of departure and pharmacokinetic models they propose to use. Quotes were taken from the US EPA (2020a):

General Assumption 1: "that a bioactive nominal *in vitro* assay concentration approximates the *in vivo* plasma concentration that would correspond to a similar effect." The EPA should weight the merits of using an interspecies uncertainty factor (UF) for pharmacodynamics prior to applying reverse dosimetry models to points of departure from NAMs relying on rat cells.
General Assumption 2: "that *in vivo* plasma concentration can be approximated based on steady-state kinetics." The EPA should assess the time to steady state for the members of this class of chemicals based on the available toxicokinetic literature, time course toxicokinetic models, and key physicochemical properties.
General Assumption 3: "that a toxicokinetic model to estimate external exposures…that may have resulted in that plasma concentration can be constructed using estimates of species-specific physiology and Phase I and Phase II enzyme-driven hepatic clearance." When species- and chemical specific metabolism rates are lacking, there are multiple approaches that could be used to generate an estimate. The EPA has chosen direct extrapolation from one species to another. At least two Panel members recommended that rat metabolism data be developed to facilitate comparisons of NAMs to rat *in vivo* data. When species and chemical specific *in vivo* or *in vitro* metabolism parameters are not available, the EPA should weigh the merits of the following alternative approaches based on the quality of the available data: 1) use a species-specific value

for a chemical deemed to be most structurally similar (e.g., similarity as identified using the EPA Chemicals Dashboard) (Lu et al., 2016); 2) use or develop a quantitative structure-activity relationship or quantitative property-property relationship for intrinsic hepatic clearance of structurally similar chemicals; 3) use a categorical approach to estimating typical rat/human clearance ratios for a class of compounds (e.g., similar to approach of Béliveau et al., 2005 for volatiles, based on hepatic Cytochrome P450 Family 2 Subfamily E Member 1 (CYP2E1) concentrations (though no consistent ratio is evident from the hepatic intrinsic clearance (Cl_{int}) values for OPs in the Supplemental Appendix for DNT-NAMs, Table 4).

Simplifying General Assumption #1. "100% bioavailability" by the oral route/portal vein. When applied to human safety assessment in IVIVE, this assumption is health protective. Empirical evidence supporting or refuting this assumption should be provided for members of the subject class of chemicals. Relying on this assumption to compare rat *in vitro* and *in vivo* effect levels has different implications—trying to perform best comparisons is different from making conservative policy assumptions. EPA should use a bioavailability predictor which would provide more relevant comparisons of *in vivo* data and predictions of external equivalents of *in vitro* effect levels.

Simplifying General Assumption #2. No extrahepatic metabolism. The only occurrence of the term "extrahepatic" in the EPA Issue Paper is at this location. EPA should summarize the potential for extrahepatic metabolism of OPs in relevant species and the impact of extrahepatic metabolism (if present) on a risk assessment.

Simplifying General Assumption #3. "Hepatic metabolism is first order (proportional to concentration) and does not saturate". This assumption is problematic if one is trying to use HTTK models for dose response assessment in rat studies, as traditional dose selection paradigms often put upper doses in the saturated realm. The EPA should evaluate the validity of this assumption for OPs in the range of NAM points of departure based on comparison.

Simplifying General Assumption #4. "Renal clearance is proportional to fraction unbound in plasma (Fup) and glomerular filtration rate (i.e. no active transport)." The EPA should evaluate available evidence supporting or refuting this assumption via literature searches for the chemical name and chemical class name and "active transport" (and other appropriate key words). If no data can be identified, it should be noted as data gap, and the implications for risk assessment delineated.

Simplifying General Assumption #5. "No biliary excretion or enterohepatic recirculation occurs." The EPA should evaluate this assumption for OPs because it is NOT a health protective assumption in the human.

A Panel member had some concerns about HTTK model performance and use of PBPK models, as discussed in the following: EPA states, "HTTK models have demonstrated reasonable accuracy" (Wambaugh et al., 2018). In the EPA Issue Paper, metrics for agreement are r^2 values. This metric was not appropriate as one can achieve high correlation (r^2) without high accuracy if all of the data are off by a similar relative amount in the same direction. The EPA should restate the model accuracy in terms of the fold error in the prediction (Wambaugh et al., 2018, Figure 9). Most chemicals were not "in the range" (i.e., within ~3x, or half an order of magnitude). The EPA should quantify the "fold error" performance of the 3-compartment HTTK model for each parameter (maximum concentration [Cmax] and area under the curve [AUC]), for all available data. The EPA should report the following statistics: % within 2-fold (World Health Organization International Programme on Chemical Safety, 2010 criterion), % within ½ log unit (3.16) (standard EPA interspecies toxicokinetic uncertainty factor), % within 1 log unit, the

range of fold error corresponding to 95% of the data, and % of data overpredicted. The EPA can then compare errors for the case study chemicals to the universe of HTTK-modeled chemicals. The EPA should avoid value judgements such as "reasonable" accuracy outside of the context of a specific risk assessment purpose, as desired/required level of accuracy differs from prioritizing functions vs. chemical specific risk assessments. At least two Panel members recommended that EPA should not generally use HTTK models for chemical-specific risk assessments.

One Panel member noted that the EPA generated comparisons of the outputs of multiple modeling assumptions which were in accordance to each other (rat vs. human intrinsic clearance). The standard of agreement that was established sets a lower expectation (agreement within ~3x or ~10x) and the numbers of comparisons lying within those ranges were not quantified. The EPA should compare HTTK model predictions to measured values instead of or in addition to their "rat to humanized rat" comparisons.

A Panel member noted that the EPA provided comparisons to PBPK models where such models were readily available. The lack of agreement between the PBPK and the HTTK models for two of three chemicals (US EPA 2020a, Figure 8) where comparisons could be made were not commensurate with their overall physiological and/or biochemical fidelity/accuracy for this class of chemicals. If additional comparisons can be made using the chlorpyrifos and diazinon PBPK models, that would add to the weight of evidence. The Agency stated that the chlorpyrifos PBPK model cannot be utilized because it was written in the PK software package acs1X (US EPA 2020a, page 62). The EPA should try to increase the number of HTTK to PBPK model comparisons. The EPA should pursue the following options: 1) Contract with people who are still using acslX for PBPK modeling. 2) Translate ACSL/acslX models to Magnolia or another platform. 3) Consider use of the diazinon PBPK model (Poet et al., 2004).

Another Panel member noted that the IVIVE approaches used for the calculation of NAM AEDs were consistent with the current state of the art for rapid evaluation of high throughput *in vitro* testing and are appropriate for use in the preliminary comparison with *in vivo* data on acetylcholinesterase inhibition described in the EPA Issue Paper. However, the same Panel member strongly recommended that, in the future, *in vitro* metabolism data be collected in the species of interest for the administered equivalent dose, rather than using available *in vitro* metabolism data from another species as a surrogate. That is, metabolism data in the rat should be used for the comparison with *in vivo* inhibition data in the rat, but metabolism data in the human should be used for calculation of Oral Equivalent Doses (OEDs) in the human. Although the comparisons of AUCs and Cmax's predicted using rat and human metabolism data suggested that they are strongly correlated, it also shows that the impact of differences in rat and human metabolism for a specific chemical can result in more than an order of magnitude difference in predicted NAM AEDs.

A different Panel member emphasized the importance of recognizing that the rapid screening HTTK IVIVE approach used for this comparison would be less appropriate for comparing alternative points of departure in a chemical-specific risk assessment. In support of a risk assessment, a more robust, quantitative IVIVE approach (Yoon et al., 2012) should be used. There are several simplifying assumptions associated with the HTTK modeling approach that were chosen to provide a conservative (health protective) bias for rapid model predictions of OEDs in the human to support early screening and prioritization. That is, the assumptions tend

to lead to a lower estimate of the OED or human equivalent concentration (HEC) AED that is associated with a given *in vitro* bioactive concentration. However, when estimating a point of departure for possible incorporation in a risk assessment, the desire would be to make assumptions that tend to lead to a more accurate estimate of the NAM AED. The comparison of the HTTK AEDs with PBPK AEDs for a few chemicals suggested that these conservative assumptions could result in substantial uncertainties in NAM AEDs. Of the simplifying assumptions used in the HTTK model, the ones that have the greatest impact on a NAM AED are:

100% Oral absorption: Pharma has developed an empirical approach for estimating absorption of drugs using data from human epithelial colorectal adenocarcinoma CACO-2 cells and *in vivo* studies, however, since it was based only on data for drugs, it was not generally applicable for environmental chemicals, which have a much broader range of physico-chemical properties.

Pre-systemic metabolism in the intestinal tissues can be ignored: This assumption is often incorrect for oral exposures because the cells lining the intestines possess significant metabolic capability. Because of the significant impact of intestinal metabolism on drug delivery, pharma routinely includes intestinal oxidative cytochrome P450 (CYP) metabolism in their IVIVE predictions by using CYP isoform abundance data in PBPK modeling platforms like SimCyp.

Metabolism in systemic tissues other than the liver can be ignored: While the liver provides the vast majority of systemic CYP metabolism capability, significant esterase metabolism capability was present in other tissues and in the plasma (particularly in the rodent). In the case of organophosphates (and many environmental esters), ignoring both plasma and intestinal esterases can seriously underestimate clearance.

Metabolic clearance is linear: This assumption is appropriate when the model is being applied at exposures that are sufficiently low to avoid saturation of metabolic processes, and is routinely applied when modeling the environmental (but not occupational) exposure levels typically anticipated for the human population or when calculating HEDs from *in vitro* bioactive concentrations.

Only the unbound fraction of the chemical in the plasma is available for metabolism in the liver (a.k.a., restrictive clearance): This assumption is incorrect in principal (because it ignores Le Chatelier's principle that an equilibrium shifts in the direction of product when product is removed), but it has nevertheless been used by pharma as a convenience. However, it has been demonstrated that the alternative assumption that all of the chemical in the blood is available for metabolism (non-restrictive clearance) results in a more accurate prediction of *in vivo* clearance (and the resulting steady-state *in vivo* concentrations) for chemicals in Toxcast (Wetmore et al., 2012), and this is the assumption that has generally been used in PBPK models for environmental contaminants and pesticides.

Renal clearance is by glomerular filtration of unbound chemical: This assumption is usually acceptable, but it is incorrect for chemicals like perfluoro-octanoic acid, that are retained by a saturable resorption process.

No biliary excretion or enterohepatic recirculation occurs: There is currently no way to address this concern in the absence of *in vivo* pharmacokinetic data, but it is not likely to be as important an assumption as the others listed above.

A Panel member recommended that these simplifying assumptions in HTTK modeling should be re-assessed to identify whether different assumption and additional chemical-specific *in vitro* data would improve the accuracy of the NAM AED estimates, and to determine potential approaches for their more accurate application in chemical-specific risk assessments. For

example, incorporation of saturable liver, intestinal and plasma metabolism would only require *in vitro* assays, and together with the use of non-restrictive clearance would greatly improve the accuracy of NAM AED predictions. The Panelist pointed out that these issues were currently under study in an ongoing research effort being conducted under the Cosmetics Europe Long Range Research Strategy, in which the US EPA ORD is participating. One of the goals of this effort is to develop internal Thresholds of Toxicological Concern for use in the safety assessment of cosmetics (Ellison et al., 2019). The European Commission's Joint Research Centre and US EPA ORD are also participating in this effort, in which the objective is to develop appropriate approaches for estimating conservative (lower) estimates of the steady-state blood concentrations associated with dosing in the animal studies. These animals studies served as the basis for the development of threshold of toxicological concern (TTCs), in order to be able to apply the TTC concept to inhalation and dermal exposures in humans. The Panelist recommended that the US EPA Office of Pesticide Programs (OPP) take advantage of ORD's involvement in this research to determine the appropriate values in improving pesticide risk assessment.

A Panel member commented that when using the highly conservative HTTK model, the NAM AEDs were substantially higher than the *in vivo* doses associated with AChE enzyme inhibition. Moreover, NAM AEDs calculated with PBPK models were even higher than those obtained with the HTTK model. As a result, supporting an inference that the effects observed in the *in vitro* assay occurred at higher concentrations than those associated with AChE inhibition. However, the same Panel member pointed out an important uncertainty that needs to be considered before drawing this conclusion: the reliance on nominal concentration to characterize cell and tissue exposures in the *in vitro* assays. The same member also noted that although chemical kinetics and free concentration have typically not been considered by the pharmaceutical ("pharma") industry when conducting *in vitro* screening of drug candidates, pharma is typically satisfied with qualitative rather than quantitative predictions. Moreover, greater challenges have been encountered with environmental compounds, primarily due to their wider range of chemical properties and more diverse array of exposure scenarios compared to drugs. The same Panel member continued to note that one of the chief limitations of *in vitro* assays identified in the evaluations was insufficient data defining the chemical domain of applicability of an assay which was intended for use with environmental chemicals. The assays had been validated using only drugs. Another Panel member commented that binding of a chemical to proteins/lipids in the media and to materials in the exposure system can be a key determinant of free concentration. Additionally, it was one of the crucial considerations for a successful IVIVE due to binding limits compounding the availability for uptake into cells and tissues *in vitro*; therefore the free concentration must be determined in order to accurately determine the *in vivo* plasma concentration that would be expected to elicit a target-tissue response similar to the cellular response in the *in vitro* assay (Birch et al., 2019; Groothuis et al., 2015, 2019; Kramer et al., 2012, 2015). The Panel member stated that the potential inaccuracy associated with the use of nominal concentration has been an "inconvenient truth" that has been resisted for several decades due to the perceived cost and time associated with the necessary analytical chemistry, but that concern is no longer justified. The same Panel member strongly recommended that the EPA take steps to assure that *in vitro* NAMs are appropriately evaluated by measurement of the free concentrations of representative chemicals spanning the chemical space intended for the assay, under standard assay conditions.

A Panel member recommended adding chlorpyrifos-methyl to the OPs included in the case study to assure that all members of this class of insecticides currently-registered in the U.S. are evaluated with this new and promising approach to understanding the potential for neurodevelopmental toxicity of these chemicals. The same Panel member also offered the following comments: The EPA Issue Paper, Figure 7 presented a comparison of NAM-generated Administered AED50s and *in vivo* data-derived Benchmark Dose (BMD) values for most of the OPs evaluated in the MEA NFA and the HCI assay. The BMD values were derived from data in rats on AChE inhibition, primarily in red blood cells (RBCs), although such data also may have been collected in plasma and brain. These data currently serve as the primary source for the derivation of guidance values (e.g., oral reference doses (RfDs)) in the human health risk assessment of OP pesticides.

A Panel member mentioned that several Panel members noted that the graphical presentations of the rat and human *in vitro* to *in vivo* extrapolation (IVIVE) comparisons were very useful in visualizing an extensive amount of information. The discussion that followed the Agency's presentation on the NAM OP case study on Day 1 of the meeting was very helpful in clarifying what initially had been characterized by several Panel members as an apples and oranges comparison. The cautionary note is that one should not demand more of this information than it can convey at the present time.

Several Panel members commented that the scientific community currently struggles with the question as to whether neurodevelopmental effects of concern may be occurring at dose levels lower than the hazard guidance values based upon cholinesterase inhibition. It was noted that in the EPA Issue Paper, Figure 7 does not provide answers to this question at this time.
A Panel member noted that the EPA's risk assessment community generally interprets the lower confidence bound on benchmark dose corresponding to an *x*% increase in response (BMDL10) as a value equivalent to the historical No-Observable-Adverse-Effect level (NOAEL) and as the Point of Departure (POD) to which uncertainty factors can be applied to derive reference values such as the Reference Dose (RfD). This is, in fact, the practice of EPA's Office of Pesticide Programs when deriving both acute and chronic oral RfDs for all of the OPs for which there were adequate data to model the dose response for cholinesterase inhibition. In a few cases where this was not possible, OPP derived acute and chronic RfDs by applying uncertainty factors to the NOAELs identified in the most relevant studies.

A Panel member noted that the EPA's ORD scientists described, in some detail, how and why they derived AC50s and AED50s, acknowledging that they should not be interpreted currently as No-effect levels. The ORD noted that they are continuing to refine the technology and to work towards the goal of being able to characterize other benchmarks that may more closely represent a no-effect level so that the comparisons can be seen as equivalent. Only then can a judgment be made on whether hazard guidance values based upon cholinesterase inhibition are protective against neurodevelopmental effects of concern.

Another Panel member offered the following limitation to the IVIVE approach: 1) In sections 2.3.6.1 and 2.3.6.2, (US EPA, 2020), a number of simplifying assumptions are declared with little or no upfront justification. In section 2.3.6.1, a general assumption of the IVIVE approach was that a bioactive nominal *in vitro* assay concentration approximates an *in vivo* plasma concentration that would correspond to a similar effect. How true would this be for compounds

that need to traverse the blood brain barrier and may encounter specific transporters? This latter consideration could have quite an impact on uptake into the target tissue. In section 2.3.6.2, a major simplifying HTTK assumption of the IVIVE is the absence of extra-hepatic metabolism. This assumption would seem to be particularly misplaced with regard to organophosphates, which have well known interactions with plasma esterases, including paraoxonase (Ceron et al., 2014) and carboxylesterase (Talcott, 1979).

Another Panel member noted that organophosphates are only one very small group of chemicals with the possibility of causing developmental neurotoxicity. As a class, however, OPs have a common mechanism of action (inhibition of acetylcholinesterase), which was why OPs were used as a test case for IVIVE. The EPA reported that the study authors noted that some endpoints in the DNT-NAM panel were not selective and that some endpoints could not be obtained for some of the test OP compounds. However, considerable work was done to calculate approximate AEDs for the majority of OP compounds tested and to compare these to benchmark doses in both human and rats. Both PBPK-derived and HTTK-derived AED's were determined. However, at the time the report was written, data were not always available so this could not always be done in both species. Where it could, the AED's in both species were approximately similar and were generally higher than benchmark doses. With caveats noted by other members of the Panel, it appeared that the IVIVE approach was reasonable, and is likely to be helpful with DNT-NAM assay data, including MEA NFA and HCI assays.

Charge Question 5. – ***Data-derived Extrapolation Factor's Using In Vitro AChE Inhibition Data***

Question 5. *In vitro* acetylcholinesterase inhibition data have been generated for rats and humans to develop interspecies and intraspecies data-derived extrapolation factors (or DDEFs) for pharmacodynamics for 16 organophosphate compounds in accordance with the EPA's 2014 *Guidance for Applying Quantitative Data to Develop DDEFs for Interspecies and Intraspecies Extrapolation*. The studies are briefly described in Section 3.2 of the EPA Issue Paper and more details can be found in MRIDs 50773501 to 50773503. Please comment on the strengths and limitations of these data. *Please include in your comments a consideration of the study design and methods, appropriateness of the selected measures, sufficiency of reporting, and robustness of the in vitro acetylcholinesterase inhibition data, including sample size.*

STRENGTHS:

Evaluation of measurable event associated with central nervous system (CNS) toxicity.
Data developed to formulate DDEF values.
Original data were made available for evaluation.
Technically reliable measures of acetylcholinesterase inhibition (AChEI) were developed.
Experimental results available from relevant test species and humans.
Both sexes were evaluated.
Effort was made to address intraspecies variability through multiple samples.
Individual samples, not pools, were evaluated for humans.
Experimental design allowed for direct comparison of results between and among species.
Effort was expended to demonstrate statistical validity of derived values.
Evaluation of findings by multiple organizations.

Sufficient data appeared to have been collected to allow for a reanalysis to address some of the weaknesses identified below.

LIMITATIONS:

Red blood cell (RBC) AChE may not fully represent brain AChE.
Some uncertainty accompanies assigning AChEI results from RBC to brain enzyme.
Animal data were not generated in the susceptible lifestage.
The number of samples appeared insufficient to characterize human variability.
The number of samples appeared insufficient to characterize the central tendency of rat data.
The demographics of the human samples does not reflect population heterogeneity with respect to (e.g.) lifestage, age, ethnicity.
Animal samples were pooled, rather than individually analyzed.
The analysis of AChEI to derive the bimolecular rate constant was not consistent with DDEF guidance, precluding comparison of candidate DDEF values across a range of concentrations.
The range of organophosphate concentrations used was not compared to concentrations expected to be attained in exposed animals (at the point of departure) or humans.
AChEI may occur at concentrations low enough to escape detection.
AChEI data were not interpreted according to DDEF guidance.
Statistical analyses:

- Uncertainties in some parameters (e.g., k' estimates) were not consistently addressed
- Statistical Analysis System datasets do not include standard error estimates for input response variable, KAPP, resulting in residual uncertainty not being fully characterized
- Robustness was complicated by applying linear and non-linear methods to a reaction that is neither first order nor second order
- Reliance on Warning indications may carry more weight than is valid
- It was not clear that sufficient statistical treatment of outliners has been applied
- Intra-person variability (variability shown by multiple measures from the same individual) seems not to be accounted for in most of the analyses
- Sufficiency of Sample Size: There are issues with coefficient of variability that impact sample size acceptability which are discussed more completely below

RECOMMENDATIONS

1) Provide increased detail in explaining why acetylcholinesterase inhibition in red blood cells is representative of acetylcholinesterase inhibition in brain tissue. This should include a description of the data and interpretations applied to acetylcholine esterase in red blood cells and brain, as well as the molecular interaction and subsequent modification leading to inhibited enzyme.

2) Develop a sample pool (not a pooled sample) of humans more reflective of the human population, to include more individuals from specific racial and ethnic backgrounds, as well as humans from the susceptible (developing) lifestage.

3) Increase the sample size in rats to include not only more samples, but samples from the susceptible life stage (developing animals).

4) Conduct or present data that demonstrates that the inhibition potential for perhaps a subset of these OP chemicals is the same/similar in brain and red blood cells. This will be necessary to develop an increased level of confidence in findings from studies of red blood cells.

5) Present findings directly in context of the EPA's Data Derived Extrapolation Factors guidance (US EPA, 2014). This should include presentation of inhibition determined at multiple concentrations, and candidate DDEF values calculated for different concentrations; as well as a comparison of the concentrations studied in vitro to the concentrations of OP attained in animal studies at the animal in vivo point of departure as well as the doses/concentrations to which humans are expected to be exposed.

6) Increase the presentation and discussion of key statistical issues identified further in this report. Address the issues related to the statistical analysis including:

- Address, discuss, analyze uncertainties in statistical parameters evaluated.
- Include standard error estimates for all input response variables.
- Fully discuss issues related to impaired robustness introduced by applying linear and non-linear methods of analyses to a reaction that is neither first order nor second order.
- Decrease reliance on Warning indicators as a criterion for exclusion of model results.
- Increase the level of clarity regarding the statistical identification and exclusion of outlier data.
- Increase the level of clarity regarding documentation and explanation that measures of intra-person variability (uncertainty produced by demonstration of different measures of inhibition from replicate samples) is consistently accounted for in all analyses.
- Include a discussion of how the coefficient of variability impacts sample size, how such data were included in decisions of optimal sample size and the impact of coefficient of variability on the level of confidence that can be placed on the results.

COMMENTS:

Several areas of the analysis received particular attention during the review meeting. These included addressing concerns regarding:

- The extent to which the data interpretation followed the EPA's guidance
- The acceptability of the number of rat and human samples
- The acceptability of the lifestage of rat samples analyzed
- The acceptability of the breadth of human population covered by the human sample set
- The extent to which the analysis focused on the bimolecular rate constant
- Choice of model selection based on software error warnings
- Conclusions reached on the basis of the statistical analyses

Comments from one Panel member focused directly on the DDEF guidance, and emphasized key points relative to the present analysis, citing the importance of DDEF sections and subsections 2, 2.2, 2.4, 4.1, 4.1.2, 4.2, 4.2.1.4, 4.2.2.1 (US EPA, 2014). The same Panel member cited the EPA DDEF guidance differentiating *in vitro* measures from *in vivo* measures and indicating, "DDEF values for toxicodynamics (TD) may also be quantified as the ratio of *in vitro* concentrations

producing the same level of response." The Panel member reviewed DDEF toxicodynamic guidance, including the evaluation the concentration-response relationships in respective species and samples to determine whether nonlinearities in the underlying data set are present and which might influence the value of the DDEF TD factor when derived at different inhibitor concentrations. In advocating a comparison of DDEF values derived at multiple concentrations, the EPA guidance intends to guard against uncertainty produced by nonlinearities in data sets, the differential presence of nonlinearities in data sets under comparison and differences in the slope of the concentration-inhibition curves describing the inhibition function in (e.g.) species under comparison. Any of these factors would produce instability in the relationship between DDEF value and the response level at which the ratio of concentrations was determined. Indeed, such nonlinearity was observed at low levels of inhibition during the same Panel member's own evaluation of data presented for bensulide oxon. The Panelist indicated that data describing the relationship between OP concentration and the degree of inhibition are presented for study samples in cells A26 through G33 of the "summary" worksheets in excel files for individual samples, and presented the results of an example DDEF raw data spreadsheet calculation from one rat and one human (bensulide oxon).

The Panelist emphasized that the EPA's DDEF method (1) specifies "concentration" as the unit of measure, (2) advocates a comparison of in vitro study concentrations to tissue concentrations attained in animal studies at the risk assessment point of departure and (3) states that a presentation of the rationale and implications for choosing the response level used as the point of extrapolation should be developed. To the extent that the derived biochemical term, *ki* (the bimolecular constant), is a single value (a point estimate derived from an amalgamation of data from multiple times and multiple concentrations) and not a concentration, and reliance on *ki* precludes a more detailed analysis of concentration-response data as required by DDEF methodology, the use of *ki* in deriving DDEF values cannot be justified – its derivation quite likely includes data from concentrations that are widely divergent from those approximating concentrations producing the level of inhibition expected at the risk assessment point of departure and, further, include data from physiologically irrelevant and unjustified OP concentrations. A Panelist included an example analysis of DDEF calculation using data from one rat and one human to emphasize these points, demonstrating the value of data already developed to serve as the basis for DDEF derivation.

The same Panelist was of the opinion that the US EPA selected RBC AChEI as a surrogate for the critical effect substantially based on expediency and cost, relative to time and conservatism of the number of animals required to supply sufficient tissue for evaluation. This contention of inappropriate application of data from these samples was supported in US EPA 2020b which states that there are differences between RBC and brain AChE in the peripheral anionic site that can alter binding, resulting in differences in binding at the catalytic site of the enzyme. This memo also identified a potential impact of yet-unstudied post-translational modifications and changes in other allosteric sites outside the catalytic site that can also influence the binding of substrates. While it is presumed that AChE activity in the RBC is a suitable representation of AChE activity in the brain, few conclusive studies are available to support this contention. Short of a reliable comparison of RBC-brain AChE activity and inhibition potential, the use of RBC AChEI as a surrogate for brain AChEI might be acceptable in some situations (i.e., where a very low and perhaps not biologically adverse level of inhibition is used to determine a risk assessment point of departure), but is unsuitable for a more complex evaluation of variability.

While the experimental design and data presentation was sufficient to demonstrate empirical findings of concentration (and time-) dependent enzyme inhibition, some discussion focused on inter- and intraspecies comparison being made at the level of a derived value (the bimolecular rate constant), rather than empirically-observed concentration-dependent inhibition data. These discussions also focused on the intent of the DDEF guidance (sections noted above) to force a concentration-based presentation of not only toxicokinetic, but also toxicodynamic comparisons.

Multiple Panel members noted that the reporting was insufficient. One Panel member observed inconsistencies across the several submitted documents. A second Panel member remarked on the use of an incorrect term, while a third Panel member observed the lack of documentation that explained what was reported in the data files. One of these Panel members also questioned the robustness of the data (covered more fully below), indicating that uncertainty exists when, for example, a pseudo first order approach was taken to estimate the *ki* parameter for a reaction that was not in the first order.

A third Panel member concluded that the biochemical method used to determine AChE inhibition was appropriate but questioned several other aspects of the study. Selection of the most appropriate model on the basis of a Warning label over an AIC value prompted the Panelist to suggest that increasing the sample size might clear the Warning. Large variability (21-fold) among humans was noted for omethoate, with within-subject variability also noted as high (approximating 3-fold). Sample size was noted as being too small for some population groups (e.g., African Americans, Hispanics) to draw any meaningful conclusions. Pooling of rat samples by sex was criticized. The Panelist noted that there was little or no interpretation of the results in the Conclusions sections of the submitted reports.

One Panel member addressed variability of *ki* values, noting a range of variability generally in agreement with what might be expected from blood samples used for clinical evaluations; this same Panel member observed that "the overall range of rate constants for each OP compound tested was relatively small." Noting that a relatively low number of rats from a single strain was used in the analysis, the Panelist indicated that the need seems to be for more rat samples, rather than more human samples. Noting the value of concentration-response data, the same Panelist indicated that AChE inhibition may actually occur at concentrations that are orders of magnitude lower than those tested, perhaps even lower than traditional limits of detection.

<u>"Representativeness" and sample size</u>:
One Panel member introduced the term "representativeness" during the meeting discussions when articulating his concern that the human study did not adequately represent the human population, either in its sample size or the nature of population differences particularly, in age or race/ethnicity. Another Panel member pointed out that race and ethnicity appeared to be conflated and that one cannot automatically assume one from designation of the other. It was also noted that neither Native Americans nor Asians were represented at all.

There was a critical flaw in both the rat and the human study in the failure to adequately examine the impact/import of age on the outcome. The US EPA 2020b presented five hypotheses related to how pharmacodynamic parameters (PDPs) associated with AChE inhibition would compare across species (rat to human) and within species (human). Two of them (#3 and #4) posit

specifically that age, gender, or disease status would have no effect, either within or across species.

At least one Panel member found the argument that age has no impact on pharmacodynamics parameters (PDP) unconvincing pre-meeting, and was not convinced otherwise as a result of discussions at the meeting.

The Health Effects Division (HED) of the Office of Pesticide Programs frequently evaluates risk for the following U.S. population subgroups: all infants (<1 year old), children 1-2, children 3-5, children 6-12, youth 13-19, adults 20-49, females 13-49, and adults 50-99 years old. Females 13-49 are singled out for several reasons: they are of child-bearing age; they are the vehicle/surrogate for prenatal exposure; they are usually considered the most sensitive subpopulation in the occupational setting. Age groupings may be adjusted somewhat on a chemical-by-chemical basis, depending upon the specific exposure scenarios presented by the use patterns of the pesticide being evaluated.

So, how does this information affect an assessment of the design of the Data-derived Extrapolation Factors (DDEF) rat and human studies? It renders them both inadequate, but in somewhat different ways. With regard to the rat study, additional groups, representing younger life stages, should be included. Selection of which age group(s) should be tested can be informed by the designs for the Acute or Repeated or Gestational Comparative Cholinesterase Assays (CCA), non-guideline studies that have been required for OP parent compounds and some metabolites. In these assays, single or 10 daily repeated gavage doses are administered to neonatal (beginning postnatal day 11) and to young adult (e.g., postnatal day 58-63) rats of both sexes or to pregnant dams on gestation days 6-20 (three or four treatment groups and a control group). In the acute study, peak cholinesterase (ChE) inhibition in plasma (sometimes), red blood cell (RBC) and brain was measured 4 or 8 hours following dosing. Most often, the acute and repeated dose studies results show that the neonatal pups are more sensitive than the young adults (that is, they show effects at lower doses). However, none of these study designs allow for parsing any differences to pharmacokinetic (PK) or pharmacodynamic (PD) factors; either or both could play a role.

During the question clarification session following the presentation to the Panel of the results of the rat and human studies, several Panel members raised the issue of lack of representation of key age groups in one or both studies. One of the agency representatives asked what impact would it have on choosing to expand the rat study (or not) if one already knew the comparative sensitivities of the young versus adults, based upon the results of the CCA studies. As pointed out above, these *in vivo* studies do not allow for parsing any differences in the PK or PD contributing factors and thus would be of little use in deriving Data-derived Extrapolation Factors (DDEFs) for specific chemicals.

Turning to the human data, an N = 18 was simply not robust enough no matter how rigorous and tortuous the statistical analysis may be. Matching up current numbers of humans evaluated against the subpopulation groups generally assessed, one sees 4 newborns, but 0 older infants in the all infants (<1 year old) category, 0 in the children 1-2 category, 0 in the children 3-5 category, 4 in the children 6-12 group, 2 in the youth 13-19 group, 6 in the adults 20-49 group, 5 in the females 13-49 category, and 2 in the adults 50-99 year old group. Children in the 3-5 and

6-12 year old groups are frequently those assessed to be at greatest risk given their dietary and behavioral habits, exposures per unit body weight and active stages of neurodevelopment. It would be particularly important to have them represented in the sampling. At least two Panel members argued that a sample size at least 60-80 individuals, appropriately spread across age, gender and race/ethnicity groups, was warranted.

A Panel member noted lack of defined section addressing the developmental maturation aspect of the studied enzyme in blood and brain and recommended inclusion of documentation of the assumptions supporting the reliance of data from RBC AChEI as representative of how each organophosphate will reach and react with brain enzyme. Such a discussion should also include whether there are known mutations that might affect AChE function.

Sufficiency of sample size:

The parameter of interest in this study was the interspecies pharmacodynamic DDEF (denoted as EF_{AD}) which is estimated by the ratio.

$$EF_{AD} = k_{i,Human} / \ k_{i\_Rat}$$

While estimated k_i values are available for individuals, EF_{AD} was estimated as the ratio of the mean of the $k_{i,Human}$ from the human samples, and the mean of the k_{i_Rat} for the rat samples.

In the EPA Issue Paper, page 72, the Panel found that Exponent "provided normal Q-Q plots of the *ki* values and the ln (*ki*) values to support rationale for assuming the *ki* values were lognormally distributed for all chemicals." The EPA Issue Paper and the EPA Coversheet and ICF Statistical Analysis assumed EF_{AD} was lognormally distributed. One Panel member ran simulation of ratios of lognormally distributed random values to show that in fact EF_{AD} was adequately characterized by a lognormal distribution.

Both of these means could be assumed to be normally distributed with variability between individuals. In addition, there is variability in estimation of the k_i for an individual (within variability) (topic of Charge Question 8). It seems that within variability is not accounted for in most of the analyses.

Sample size can be estimated a number of ways, depending on how the precision target is specified.

The *Exponent sample size calculation memo* specifies a sample size determination approach that is based on a test the coefficient of variation, CV, defined as CV = s/m where one assumes that m = true (EF_{AD}) and s = standard deviation (EF_{AD}). The CV test is described in Banik et al. (2012). The test hypothesis is

H_0: CV =CV_0 versus H_a: CV > CV_0

The test statistic used is from Miller (1991) and depends on an assumption of normality and the asymptotic distribution of the sample CV. Let CVe be the estimated CV.

Then the test statistic M defined as $(CVe - CV_0)/S_{CVe}$ has a standard normal distribution where S_{CVe} is defined as the sqrt$\{(CVe^4 + 0.5CVe^2)/n\}$.

This test has the benefit of not having to specify a variance term but to rely only on the estimated value of the CV and the target difference, d_{cv}, between the observed and expected CV. Note that the Banik et al. (2012) paper also describes a test statistic by Sharma and Krishna (1994) that is more robust to the assumption of normality and which can more easily be solved directly for n given d_{cv} and the given Type I error rate, a.

Exponent in their sample size justification, set the sample size at n=18, and computed the value of ECV that would be statistically significant at a = 0.05 for a range of potential true CV values [5%, 10%, 20% and 30%] and with this computed the corresponding effective detectable difference, d_{cv}, of true population variability considered as a percent.

The computations in the Exponent sample size calculation memo were able to be confirmed. Hence a sample size of 18 is reasonable if an accuracy of ±2% is acceptable if the true CV is about 5, or ± 14% if the true CV is 30%.

Note that the Miller (1991) test statistics can be solved for *n* similar to what is done for the sample size based on specification of relative error. In this case, with type I error a specified and type II error b specified and prespecified CV target of d_{cv}, the sample size is:

$$n = (Z_a + Z_b)^2 [CVe^4 + 0.5\ CVe^2]/\ d_{cv}^2$$

For a true CV of 30%, and a difference of dcv = 14%, Type I error of 0.05, Type II error of .3, the computed n is 79 (not 18). This suggested that the power of a test to detect a 14% difference between the observed and expected CV when the true CV is 30%. at a type I error of 0.05 must be quite low (actually somewhere around power = 0.1 meaning a 0.9 probability of experiencing a type II error).

Charge Question 6. – ***Data-derived Extrapolation Factor's Using In Vitro AChE Inhibition Data***

Question 6. Given the structure of correlated data, nonlinear mixed-effects models were used to analyze the *in vitro* inhibition data in order to calculate the interspecies and intraspecies pharmacodynamic DDEFs as described in Section 3.2 of the Agency's Issue Paper and MRID 51182301. The ratios of the biomolecular rate constants between species or subpopulation were estimated from the nonlinear mixed-effects models, which are reported in Section 3.3 of the Agency's Issue Paper and MRID number 51182301. For a number of chemical-specific datasets analyzed by Exponent, the fitted non-linear mixed model generated warning statements due to a full rank final Hessian matrix. Additionally, for several of the chemical-specific datasets analyzed, visual evaluation of diagnostic plots revealed severe outliers or a severe imbalance in the distribution of residuals, leading to questionable model fit. In an attempt to resolve the warning statements and outlier issues,

the EPA consulted with its statistical contractor at ICF, which submitted a supplemental analysis (see the EPA Coversheet and ICF Statistical Analysis).

a. Please comment on the methods or techniques employed by Exponent using the nonlinear mixed-effects models.

b. Please comment on any concerns associated with the warning statements and model-fit issues. Taking into consideration the supplemental ICF analysis to address these issues, suggest, if necessary, other methods or techniques that could be suggested for addressing such warning statements and model-fit issues.

Methods used in the nonlinear mixed-effects models:

The approach to model-fitting is described on pages 11-13 of Higgins et al., 2020. The Panel concluded that the approach to model-fitting described in the Agency Issue Paper (US EPA 2020a) represents current good statistical practice and seems well thought out. Use of AIC is a typical measure used to choose among different models. Graphical diagnostics are used extensively to supplement AIC in final model selection.

The SAS NLMIXED procedure is used throughout this analysis to fit these models. Default values are chosen for all NLMIXED statement options except for some scenarios where specific options are reset to help improve convergence.

The SAS code aligns with the description provided in the Agency Issue Paper (US EPA 2020a) but includes few comment statements other than indicating which scenario is being addressed and which model is being fit. Justifications are not provided for why default model fitting options are changed for the model being fit. The logic and/or analysis approach is not presented or documented but has to be inferred by looking at which fit options have been changed from the defaults.

One Panel member referred back to the discussion of the k_{app} response variable in a previous question and recommended that the analysis start with the raw measurement that are output from laboratory experiments. This allows estimation and propagation of uncertainty in the k_{app} slope values through to the final DDEF estimates.

Recommendation: In estimating k_i, the analysis should be performed on the experimental time course measurements rather than on k_{app} values derived from regressions on individual time course measurements.

Concerns associated with model-fit issues:

The methodology described in the Agency Issue Paper (US EPA 2020a) uses the entire set of k_{app} (derived) values from individual humans and pooled rat samples to produce k_i estimates with preferred statistical properties as well as pooled estimates of residual variability. This approach is better than fitting the nonlinear model separately to the individual samples or simply using a

(fixed effects) nonlinear regression approach. Proc NLMIXED incorporates empirical Bayes estimation methodology to produce estimates for the k_i. This method produces quite good estimates for the mean k_i for humans and for rats, but the individual k_i estimates have the property that extreme values are closer to the mean than would be the case with traditional nonlinear regression (e.g. are "shrinkage estimates"). The ramifications of using shrinkage estimates when estimating the upper 95th percentile of the population, and with this statistic also estimating the ratio that makes up the intra-species DDEF are not discussed.

One Panel member commented that the issue with using an empirical Bayes approach is that the prior is constructed from the data, vs a truly Bayesian approach in which the priors are specified without consideration of the data. So, in an empirical Bayesian approach the data are used twice: to specify the mean/center of the prior (sometimes maybe even the variance) and to derive the posterior. A truly Bayesian approach would build the prior based on information external to the data. It is concerning that by using an empirical approach not only are the estimates shrunk towards a mean, but there is no opportunity to combine the external information represented by a true prior and the information available in the specific data under consideration. That said, empirical Bayesian estimators are used often. The EPA should investigate other approaches, such as using true Bayesian priors that may facilitate better quantification of uncertainty.

Recommendation: Consider assigning true Bayesian priors to better quantify the uncertainty in k_i estimates.

The model-fit issue is discussed on page 14 of MRID 51182301 (2020).

> "Model warnings were generated in some cases due to a full rank final Hessian matrix (SAS warning: "WARNING: The final Hessian matrix is full rank but has at least one negative eigenvalue. Second-order optimality condition violated"). This warning occurs when the model does not fully converge, which adds uncertainty to the estimates. If the final model selected based on AIC value had a warning statement but the alternative final model did not have a warning statement, the alternative model without the warning was selected as the final model. If both final models 2 and 3 had SAS warning statements (i.e., model fit issue), the final model was selected based on a smaller AIC value, but the results of the selected final model should be interpreted with caution."

The approach to handling the SAS warning outlined in the quote above is reasonable and represents good statistical practice. But SAS© offers some suggestions for what to do when the model does not fully converge. Since Proc NLMIXED is strongly related to Proc MIXED, some of the many points discussed in the SAS Proc MIXED manual[1] on page 6167 should be considered. It is clear that the ICF analysts incorporated several of the suggested "fixes" in their analyses, for example the use of parameter scaling, setting the max iterations higher, and the use of ridge scaling of the likelihood surface.

It is not clear that other potentially useful approaches were attempted. Particularly of note is the statement in the SAS MIXED manual that "A nonpositive definite Hessian matrix can indicate a

[1] https://support.sas.com/documentation/onlinedoc/stat/141/mixed.pdf

surface saddle point or linear dependencies among the parameters." (page 6168 of the online SAS MIXED manual[1]).

The potential for linear dependences among the parameters suggested that the model may be formulated in a way that makes it difficult to find the maximum of the likelihood surface, or that there is just not enough data to estimate the particular covariance structure selected. This latter should be examined more closely.

For example, in the fit of all the data for the chemical compound Terbufos oxon sulfoxide the final model chosen is Model 3, but the final covariance estimates are quite close to the initial values and standard errors for these parameter estimate are not provided. This suggested that the model may have been over specified and potentially Model 2 should have been considered. The analysis protocol tells us that that Model 2 was considered. The fact that it is not the final model suggests that it also had an issue with its fit and it may also have been over specified. Model 2 without the covariance term and Model 3 without covariance terms should have been fit and compared to the two models with covariance terms. In the model fit for other scenarios, the final variance and covariance terms are not statistically different from zero. This is another situation where a reduction in the covariance structure should be examined.

In many scenarios, not enough variation in the observations above the "residual" variation may be available to facilitate also estimating the covariance parameters. Here again, stepping down the complexity of the covariance matrix is a reasonable next step. That is, entertain a model where the term for the covariance between A and B is assumed known and equal to zero.

Recommendation: For scenarios where the initial estimate for the covariance between A and B is close to or equal to zero, consider fitting a model where the assumed covariance is zero.

Reparameterization of the covariance matrix is another option not examined. Rather than specify the covariance directly (the *grab*, *grhb* terms), the covariance can be specified as the correlation times the standard deviations, that is *grab = corrab * sqrt(gra*grb)*, and the model asked to estimate the correlation. Another approach is to parameterize the variance terms on a log scale in which case the final values can never be at or below zero. That is, setting for example *var(rra) = exp(gra).* The final estimate may still be quite small when all is said and done, but the final estimate will not be zero.

The SAS warning of non-convergence, which the Panel recognizes occurs seldom in this analysis, suggested the potential that the final estimates are derived not for a global maximum of the likelihood surface but for a saddle point in the likelihood surface. Evidence for this can be developed by starting the model fitting process from different initial parameter values. Varying the starting values may allow the search algorithm to eventually escape the saddle point to find the global maximum. If the final estimates are the same as before it may still be a saddle point. This is a good next step for those scenarios where the final covariance parameter estimates are not vastly different from the initial input values.

Note that SAS Proc NLMIXED has many options that allow the analyst to exert fine control of many aspects of the likelihood search process. Some that might be explored include DIAHES (a

SAS control parameter)– using only the diagonal of the Hessian matrix and RESTART=*i* – specifying that the search direction be reassessed after *i* iterations. Many other options should be explored before accepting the estimates from questionable fits.

There is no guarantee that any of the models considered for a scenario will be adequate. This is particularly the case when the data may be too sparse to adequately estimate model parameters. This issue is also discussed in Questions 5 and 7 where the impact of dividing the full data set into smaller groups for the intra-species analysis are considered.

The issue of severe outliers is more difficult to assess. In Proc NLMIXED, predicted values are computed using the parameter estimates and empirical Bayes estimates of the random effects. Standard errors of prediction are also computed and output when requested in the PREDICT OUT = statement. Since the details of the macro used to produce the residual plots was not provided in the supplemental materials, it is assumed that the residuals examined in the outlier analysis are simply the difference between the observed and predicted values of PRED (a SAS control parameter). An alternative would be to use the standard errors of the predicted values to "standardize" the residuals before being assessed. It may be that those residuals that appear outliers are actually from predictions with high uncertainty. The ICF reanalysis computed and plotted these standardized residuals to demonstrate that this in fact was the situation, with many of the residuals originally identified as outliers.

For some scenarios the outlier issue reflects the fact that the underlying data do not fully line up with the assumptions of the model. The model assumes, for example, that the phosphorylation constant {A} and the dissociation constant {B} are assumed (bivariate) normally distributed. Consider, for example, the individual sample curves provided in the supplemental files of the model fits for the chemical Naled for the rat/human scenario. The rat individual sample curves show two "groups" of responses, two in one group (high A, low B), four in the other group (low A, high B). The human sample curves display a longer low-end tail for the upper asymptotes (the estimates of A) that suggest a more lognormal shape to their distribution. This finding supports the need to consider correlation between A and B in the model although estimating a non-zero value for the correlation is not always guaranteed. While the nonlinear mixed effects regression model is likely to do a good job of estimating the parameters, there is no guarantee that the estimates for A and B effects will truly be normally distributed which may be reflected in plots of the A and B effects as well as in the estimated residuals.

As described in a similar discussion in charge question 5, a Panel member noted that the term "biomolecular rate constant," is a typo seen on line one of Section 3.2.1 in the EPA Issue Paper as well as in other places in the earlier version of the Charge Questions, also shows up on line 4 of this Charge Question. MRID 50773501 (MSU 2018) and MRID 50773503 (Chambers et al., 2018) also reference the term, "bimolecular rate constant". Because this constant plays a central role in the risk assessment, it is recommended that EPA discuss the meaning and significance of this constant when it first appears in Section 3.2.1. Inserting a sentence with references inserted on line 3 under Section 3.2.1 before the word "Briefly" as demonstrated below would help readers of the Issue Paper (US EPA 2020a) better appreciate the importance of this constant.

> "The bimolecular rate constant k_i, a measure of the inhibitory power of an organophosphate and comprised of both the binding affinity to the active site and rate

of phosphorylation (Coban et al., 2016), was originally derived and discussed in Main (1964)."

In Sections 3.2.1 and 3.2.2 of the EPA Issue Paper (US EPA 2020a), three sets of constants, the bimolecular rate constant, the phosphorylation constant, and the dissociation constant, are discussed. Across the Mississippi State study reports these three constants are designated, respectively, as ki, kp, and KI. In Section 3.2.2, the SAS analyses uses KAPP, A, and B. The graph of the Hyperbolic Plot uses k', k2, and kd, respectively. Without further clarification and explanation, the multiple terms for the same quantity is very confusing to readers. The EPA should harmonize notation for this constant in these sections to reduce reader confusion.

Charge Question 7. – ***Data-derived Extrapolation Factor's Using In Vitro AChE Inhibition Data***

Question 7. For the intra-species analyses, Exponent conducted stratified analyses, where the 18 human samples were subset into smaller groups to estimate the bimolecular rate constant ratios for these subgroups as described briefly in Sections 3.2 and 3.3 of the Agency's Issue Paper, with more details provided in MRID 51182301. EPA has concerns about the reliability of these stratified analyses due to the small sample sizes of the subgroups, as well as concerns with warning statements and outliers. EPA's statistical contractor, ICF, provided a supplemental analysis to address the warning statement and outlier issues (see EPA Coversheet and ICF Statistical Analysis). *Please comment on these intraspecies analyses performed by Exponent and their utility to evaluate intraspecies human variability in response to organophosphate exposure taking into consideration the sample sizes and the supplemental ICF analysis.*

The document "*Guidance for applying quantitative data to develop data-derived extrapolation factors for interspecies and intraspecies extrapolation"* (US EPA, 2014) indicates on page 33 two approaches to characterize intra-human variability: a first approach based on a unimodal analysis, and a second approach based on a bimodal analysis. A unimodal analysis was performed when it was not possible to identify a priori a subpopulation of sensitive individuals on the basis of physiological, biochemical or life stage-attributes. On the other hand, a bimodal analysis was used when a subpopulation of sensitive individuals can be identified. In this latter case, to derive an extrapolation factor, it was necessary to compare the central tendency in the dose metric for the general population with the central tendency in the dose metric for the subpopulation of sensitive individuals.

To quantify intrahuman variability in the bimolecular rate constant k_i, Exponent performed both types of analyses, unimodal and bimodal.

In providing an answer to this charge question, the Panel discussed both analyses, however as the charge question refers more specifically to stratified analyses, in their answer, Panel members focused particularly on the bimodal analyses.

All Panel members shared concerns similar to those expressed by the Agency regarding the reliability of the analyses' results due to the small sample sizes. Multiple Panel members pointed out the limited availability of data for certain ethnic and racial groups, and reiterated the absence

of subjects of Asian descent in the dataset despite a probably adequate sample size for Caucasian subjects. All Panel members recognized the repercussions of the small sample size in terms of statistical inference and model fitting, including convergence problems, standard errors estimated to be equal to zero, and extreme outliers (see the range of biomolecular constant values k_i values for malaoxon and omethoate). In addition, while Panel members recognized and valued the supplemental efforts of the statistical contractor, ICF, to address some of the model fitting issues, several members of the Panel agreed with the statement that ``[..] these subpopulation analyses need to be interpreted with caution [..]'' (Higgins et al., 2020, page 8).

Multiple Panel members recommended that more human blood samples be added to the already collected blood samples in an effort to increase and more effectively characterize human sample variability. The same Panel members indicated the fact that ongoing work in this direction is slated to occur at Mississippi State University under the sponsorship of a consortium of companies, as indicated in the document "*Supplemental statistical analysis of Organophosphorus (OP) pesticides in vitro inhibition study*" by Exponent, 2020 (page 35).

On the topic of samples representativeness of the human population, one Panel member offered the following comparison between the demographics distribution in the human samples and the demographics of the US general population: while adults 16-60 years of age make up about 50% of the human sample, they account for 59% of the US general population, thus highlighting the under-representation of adults in the sample. On the opposite side, the same Panel member noted an over-representation in the sample of juveniles (samples of 10-13 years of age) and infant (cord blood samples): whereas they constitute 28% (for juveniles) and 22% of the sample (cord blood), respectively, only 13% and 0.06% of the US population has an age between 10 and 19 years of age and between 0 and 5 years of age. The Panel member estimated the US general population demographics break down from the one year estimates derived from the American Community Survey of the US Census Bureau, and derived the expected percentage of the US population that was less than 6-months old as one tenth of the estimated number of US children below 5-years of age. The same Panel member noticed no significant discrepancies between the sample and the US general population in terms of gender and race: 44% of the blood samples belong to males versus 49% of the US general population was male; analogously, the percentage of white subjects in the sample was 72%, versus a proportion of 76% for the US general population.

Because of the over-representation of children and cord blood in the sample, the abovementioned Panel member argued that the distribution of bimolecular rate constants k_i would be more representative of children and cord blood k_i values than the general US population, impacting measures of central tendency as well as measures of variability, such as the standard deviation. Particularly, the Panel member observed that if the bimolecular rate constant k_i tends to be smaller for children and cord blood, then the mean and standard deviation of the sample k_i 's would also be smaller than expected. The same Panel member offered a recommendation to account for the disproportionate representation of juvenile and cord blood in the estimation of the Data-Derived Extrapolation Factor (DDEF). Specifically, the Panel member recommended either using a resampling approach with sampling weights that reflect the demographic distribution in the US general population or simply a computation of the weighted geometric mean and corresponding standard deviation. The Panel member noted that in applying the resampling strategy on the data provided, no significant difference was noted in terms of DDEF values, very

likely due to the fact that the bimolecular rate constants k_i for juveniles and cord blood in the sample tend to be very similar to those of adults.

Two Panel members observed a large range of variability in bimolecular rate constant k_i across organophosphate compounds, and another Panel member recommended that insights and more documentation regarding the extent of the difference be reported in documentation.
A Panel member noted particularly a large variability in terms of bimolecular rate constant k_i for Naled. The same Panel member remarked that some variability in the data relative to Naled could be explained by differences in solution and water content in the DMSO used to aliquot Naled. This, in turn could affect stability of storage or Naled's activation once in aqueous suspension. The Panel member also noted that some of the variability in the bimolecular rate constant k_i could be ascribed to the fact that volatile trifluoromethyl ketones are mobile on a 96-well plate, causing inhibition in adjacent wells, as also observed by Camerino et al. (2015).

Although not discussed in the charge question, some Panel members offered comments regarding the estimation of intrahuman variability performed by Exponent using a unimodal analysis approach, where the intraspecies extrapolation human toxicodynamic (TD) factor was derived as the ratio of the 97.5th, respectively, 90th, 95th, 99th percentile and the mean bimolecular rate constant in the general population. A Panel member indicated difficulties understanding how the upper percentiles for the lognormally distributed bimolecular rate constant k_i values were computed. The same Panel member also indicated an inability to replicate the calculation used to derive the Data-Derived Extrapolation Factors presented in Table 16 of the EPA Issue Paper (US EPA 2020a). The Panel member recommended that a different approach be used to estimate the bimolecular rate constant values, k_i if they are indeed assumed to be lognormally distributed: in particular, the Panel member referred the Agency and Exponent to the estimation procedures outlined in chapter 13 of Gilbert (1987).

A second Panel member elaborated that while the Panel member understands conceptually the approach undertaken to derive the intraspecies extrapolation human toxicodynamic (TD) factor, the approach currently used by Exponent does not reflect the fact that the values of the bimolecular rate constant *ki* are not known for each human, and that the estimates obtained from the PROC NLMIXED output are just estimates. The same Panel member recommended that this uncertainty was quantified and accounted for, through the use of, for example, a bootstrap approach.

Charge Question 8. – ***Data-derived Extrapolation Factor's Using In Vitro AChE Inhibition Data***

Question 8. For intraspecies analyses, a limited subset of chemicals had three replicate analytical results on each of the four sources of human samples. The results from these analyses were used by Exponent to characterize the total variability of the estimates in terms of experimental variability and subject variability as described briefly in Sections 3.2 and 3.3 of the Agency Issue Paper with more details provided in MRID 51182301. The results were not consistent across the

chemicals, ranging from 3% to 84% of the total variability due to differences in the replicate analyses.

a. Please comment on the utility of these analyses to characterize human variability in response to organophosphate exposure.

b. If there is utility in generating these data for additional OPs, please provide any suggestions for improving the design and conduct of the study.

Members of the panel recognized the effort undertaken by Exponent to gain a better understanding about the sources of variation in the human bimolecular rate constant k_i. The analyses performed by Exponent using replicate data were conducted with the goal of partitioning the overall variability of the bimolecular rate constant k_i into experimental and intra-human variability.

In assessing the utility of these additional analyses, some members of the Panel noted that even though replicate data was available, it was only available for four subjects and three chemicals. Although a Panel member noted that the data used in this set of analyses represent an improvement over previous efforts, particularly when considering the entire sample of 18 subjects, consisting of subjects of different age, gender and ethnicity, two Panel members remarked that 4 subjects are not enough to characterize the variability among humans nor to draw definitive conclusions. These two Panel members deemed the analyses performed using replicate data of limited utility.

Another Panel member found the replicate data analyses useful, particularly if considered as the beginning of a New Approach Methodology, where new data can be incorporated sequentially within each new analysis in a Bayesian framework implemented via Markov Chain Monte Carlo. As an example, the Panel member recommended the incorporation of other human data collected in similar experiments by Dr. Chamber's group. The same Panel member also envisioned the possibility of creating new synthetic data through a Monte Carlo approach, sampling data values from the estimated probability distributions of already existing human data.

Other Panel members indicated that even though, from a conceptual point of view, using replicate data to understand the amount of experimental and intra-human variability in the bimolecular rate constant k_i represented the right approach and it could be potentially useful, the implementation of the repeated analyses had some weaknesses.

As a first point of concern, one Panel member noted a difference in the procedures used to derive the bimolecular rate constants k_i used in the replicate analyses versus the procedure implemented to derive the k_i used in the interspecies and intraspecies data analyses. Specifically, the same Panel member indicated that while in the document Higgins et al. (2020), it was stated that the bimolecular rate constants k_i in the replicate analyses were estimated by fitting a hyperbolic model to the AChE phosphorylation rate data available for each individual sample (Exponent, 2020; page 75), in the interspecies and intraspecies analyses the bimolecular rate constants were estimated by fitting a joint non-linear mixed model to both human and rat data using PROC NLMIXED in SAS (Higgins et al., 2020; pages 11-15). Due to differences in procedures, the

Panel member recommended care in comparing estimates of variability across the two set of analyses for each of the three organophosphates.

Having estimated the bimolecular rate constants k_i for each human subject with replicate data, Exponent fitted linear mixed models to the bimolecular rate constants k_i for each organophosphate separately with a random effect for the human subject, thus partitioning the variance into "between subject variance" or human variability, and "within subject variance" or experimental variability. From the latter two variances, Exponent derived the IntraClass Correlation (ICC), that is, the ratio of the human variability to the total variance. Even though Panel members considered the IntraClass Correlation the right metric to characterize human variability in response to organophosphate exposure, some Panel members remarked that due to the small sample size, the tests performed were underpowered.

As a second concern, one Panel member noted that the procedure used to derive coverage ratios for the bimolecular rate constants k_i, which relies on calculating upper percentile values in the distribution of the k_i's, was based on asymptotic considerations, all derived under the assumption of large sample sizes. Due to the extremely small sample size (n=4) used in the replicate analyses, the Panel recommended that alternative approaches be employed to derive percentiles of the distribution of the k_i's, such as, for example, the approach presented in Gilbert (1987). As done in reference to other charge questions, several Panel members highlighted the problem of incorrect uncertainty quantification in the replicate analyses: as in other statistical analyses, the bimolecular rate constants k_i are used in the replicate analyses as data and not as point estimates, as they truly are.

Finally, multiple Panel members found the interpretation of the results presented in Higgins et al. (2020) quite confusing due to contradictory conclusions reported in the document. Specifically, Panel members observed how in page 34 of the document, it was claimed that "[..] for omethoate and phosmet oxon, the ICC was small (values respectively of 20% and 3%), indicating large experimental variation relative to the total variability…." whereas, few lines later, the document states that "The large Naled ICC (value of 0.84) suggests that the experimental variability is a large contributor to the overall variability.". Various Panel members observed that both statements cannot be both simultaneously correct, and that the large Naled ICC suggested that the between individual variability is a large contributor to the overall variability.

Multiple Panel members also noted that the EPA Issue Paper "*Use of New Approach Methodologies to Derive Extrapolation Factors and Evaluate Developmental Neurotoxicity for Human Health Risk Assessment*" (US EPA, 2020a) has an incorrect statement in the paragraph just before Table 17 (page 75). The statement reads: "[..] Whereas, for phosmet oxon, only 3% of the total variability was due to differences in the 3 replicate analyses of the blood sample; 97% of the observed variability was due to the differences between human subjects." As the IntraClass Correlation for phosmet oxon was reported equal to 3%, Panel members believe that 97% of total variability was due to differences in the 3 replicate analyses.

A Panel member pinpointed the differences between organophosphates (OP) in terms of human variability. Although some of the differences between organophosphates can be ascribed to sampling and methodological variability more than actual human variability according to what was stated in Higgins et al. (2020), due to the extent of the differences between

organophosphates, the same Panel member recommended that organophosphates are considered on an individual-basis and not combined into a single, joint group.

In conclusion, in reviewing the complete set of results from the replicate analyses for 3 organophosphates, some Panel members noted that within human variability was indeed important for two out of the three chemicals examined pointing to the importance of estimating within sample/experimental variability. This point was also highlighted by more recent analyses performed by Exponent on additional organophosphates.

Reviewing the replicate data analyses discussed in Charge Question 8a, several Panel members observed that considering and characterizing both human and experimental variability are important for the estimation of intraspecies pharmacodynamic extrapolation factor (EF_{HD}). Thus, various Panel members expressed support for the generation of additional replicate data for other organophosphates. In designing these additional set of replicate data analyses, Panel members provided various suggestions. All Panel members recommended a larger sample size, with more human samples chosen so that their demographic characteristics are representative of the US general population. A Panel member recommended an approach to determine the needed sample size in situations, like the one encountered here, where no knowledge of the true underlying variability was available. The same Panel member recommended a two-stage approach, called "sample size re-estimation" where an initial, typically small sample is used in a first intermediary study, from which an estimate of variance was obtained. With knowledge of the variance, in the sample-size re-estimation approach, a final sample size was derived so to achieve the desired power. The Panel member observed that a two-stage sample size re-estimation approach could be used in this context, however the Panel member warned of the possibility of oversampling. At the same time, the Panel member noted that oversampling would be easily addressed in this situation.

Another Panel member recommended that future studies pay additional care and consideration to the chemical properties of the test substances, purity of the samples, as well as possible substance mobility on 96-well plates.

Two Panel members recommended that a more accurate accounting for uncertainty be carried out in additional replicate analyses. In particular, one Panel member recommended that the uncertainty associated with extrapolating responses occurring at a cellular level to responses occurring at the whole biological system level be correctly quantified, rather than ignored without providing any rationale for it. The same Panel member also observed that responses of full biological systems might be different between the very young and the elderly, as can be the case when considering neurological systems.

A Panel member recommended that the uncertainty in the bimolecular rate constant values, k_i be properly accounted for, by acknowledging the fact that the latter are estimates and not actual data values. The same Panel member recommended a multilevel modeling approach where the raw subject data on AChE phosphorylation rate for each human blood samples are used in the analysis rather than the point estimates.

References

Banik S, Kibria BMG, Sharma D. 2012. Testing the Population Coefficient of Variation. J. of Modern Applied Statistical Methods. 11(2): 325-335.

Béliveau M, Lipscomb J, Tardif R, Krishnan K. 2005. Quantitative structure-property relationships for interspecies extrapolation of the inhalation pharmacokinetics of organic chemicals. Chem Res Toxicol. 18(3):475-485. doi:10.1021/tx049722k.

Birch H, Kramer NI, Mayer P. Time-Resolved Freely Dissolved Concentrations of Semivolatile and Hydrophobic Test Chemicals in *In Vitro* Assays-Measuring High Losses and Crossover by Headspace Solid-Phase Microextraction. Chem Res Toxicol. 2019;32(9):1780-1790. doi:10.1021/acs.chemrestox.9b00133.

Bridges DC, Tovar KR, Wu B, Hansma PK, Kosik KS. 2018. MEA Viewer: A high-performance interactive application for visualizing electrophysiological data. PLoS One. Feb 9;13(2):e0192477. doi: 10.1371/journal.pone.0192477.

Buccino AP, Hagen E, Einevoll GT, Hafliger PD, Cauwenbergh G. 2018. Independent Component Analysis for Fully Automated Multi-Electrode Array Spike Sorting. Annu Int Conf IEEE Eng Med Biol Soc. Jul;2018:2627-2630. doi: 10.1109/EMBC.2018.8512788.

Ceron JJ, Tecles F, Tvarijonaviciute A. 2014. Serum paraoxonase 1 (PON1) measurement: an update. BMC Vet. Res. 10: 74. doi.org/10.1186/1746-6148-10-74.

Camerino E, Wong DM, Tong F, Körber F, Gross AD, Islam R, Viayna E, Mutunga JM, Li J, Totrov MM, Bloomquist JR, Carlier PR. 2015. Difluoromethyl ketones: Potent inhibitors of wild type and carbamate-insensitive G119S mutant Anopheles gambiae acetylcholinesterase. Bioorganic & Medicinal Chemistry Letters 25(20): 4405-11.

Chambers JE. Meek E. and Crow JA. 2018. Inhibition Kinetics of Bensulide Oxon and Phosmet Oxon on Human and Rat Erythrocyte Acetylcholinesterase. Mississippi State University, Mississippi State, MS. US EPA MRID 50773503.

Cobana AR, Carra RL, Chambers HW, Willeford KO, Chambers JE. 2016. Comparison of Inhibition Kinetics of Several Organophosphates, including Some Nerve Agent Surrogates, using Human Erythrocyte and Rat and Mouse Brain Acetylcholinesterase. Toxicol. Lett. 248:39-45.

Corlew R, Bosma MM, Moody WJ. 2004. Spontaneous, synchronous electrical activity in neonatal mouse cortical neurones. J Physiol. 560(Pt 2):377-390. doi:10.1113/jphysiol.2004.071621.

Ellison CA, Blackburn KL, Carmichael PL, Clewell HJ 3rd, Cronin MTD, Desprez B, Escher SE, Ferguson SS, Grégoire S, Hewitt NJ, Hollnagel HM, Klaric M, Patel A, Salhi S, Schepky A, Schmitt BG, Wambaugh JF, Worth A. 2019. Challenges in working towards an internal threshold of toxicological concern (iTTC) for use in the safety assessment of cosmetics:

Discussions from the Cosmetics Europe iTTC Working Group workshop. Regul Toxicol Pharmacol. 103:63-72.

Exponent. 2020. Supplemental statistical analysis of Organophosphorus (OP) pesticides *in vitro* inhibition study. Supplement document MRID: 51182301.

Gilbert RO. 1987. Statistical Methods for Environmental Pollution Monitoring. Von Nostrand Reinhold. New York.

Groothuis FA, Heringa MB, Nicol B, Hermens JL, Blaauboer BJ, Kramer NI. Dose metric considerations in *in vitro* assays to improve quantitative *in vitro-in vivo* dose extrapolations. Toxicology. 2015;332:30-40. doi:10.1016/j.tox.2013.08.012.

Groothuis FA, Timmer N, Opsahl E, et al. Influence of *in Vitro* Assay Setup on the Apparent Cytotoxic Potency of Benzalkonium Chlorides. Chem Res Toxicol. 2019;32(6):1103-1114. doi:10.1021/acs.chemrestox.8b00412.

Higgins K, Barraj L, and Reiss R. 2020. Supplemental Statistical Analysis of Organophosphorus (OP) Pesticides *In Vitro* Inhibition Study. Exponent Report ID: 1508118.000 - 6891. Washington DC. (MRID 51182301).

Howard AS, Bucelli R, Jett DA, Bruun D, Yang D, Lein PJ. 2005. Chlorpyrifos exerts opposing effects on axonal and dendritic growth in primary neuronal cultures. Toxicol Appl Pharmacol . 207(2):112-24.

Kramer NI, Di Consiglio E, Blaauboer BJ, Testai E. Biokinetics in repeated-dosing *in vitro* drug toxicity studies. Toxicol *In Vitro*. 2015;30(1 Pt A):217-224. doi:10.1016/j.tiv.2015.09.005.

Kramer NI, Krismartina M, Rico-Rico A, Blaauboer BJ, Hermens JL. Quantifying processes determining the free concentration of phenanthrene in Basal cytotoxicity assays. Chem Res Toxicol. 2012;25(2):436-445. doi:10.1021/tx200479k.

Lefebvre B, Yger P, Marre O. 2016. Recent progress in multi-electrode spike sorting methods. J Physiol Paris. 110(4 Pt A):327-335. doi: 10.1016/j.jphysparis.2017.02.005.

Lu J, Goldsmith MR, Grulke CM, Chang DT, Brooks RD, Leonard JA, Phillips MB, Hypes ED, Fair MJ, Tornero-Velez R, Johnson J, Dary CC, Tan YM. 2016. Developing a physiologically-based pharmacokinetic model knowledgebase in support of provisional model construction. PLoS Comput Biol. 12(2):e1004495. doi: 10.1371/journal.pcbi.1004495.

Main, AR 1964. Affinity and phosphorylation constants for the inhibition of esterases by organophosphates. Science 144:992-993.

McGrayne SB. 2011 The Theory That Would Not Die, How Bayes' Rule Crack the Enigma Code, Hunted Down Russian Submarines & Emerged Triumphant from Two Centuries of Controversy. Yale University Press.

Miller EG. 1991. Asymptotic test statistics for coefficient of variation. Communications in Statistics-Theory & Methods. 20: 3351-3363.

MSU. [Mississippi State University on behalf of FMC Corporation] 2018. Inhibition Kinetics of Malaoxon and Omethoate on Human and Rat Erythrocyte Acetylcholinesterase. Mississippi State University, Mississippi State, MS. US EPA MRID 50773501.

Poet TS, Kousba AA, Dennison SL, Timchalk C. 2004. Physiologically based pharmacokinetic/pharmacodynamic model for the organophosphorus pesticide diazinon. Neurotoxicology. 25(6):1013-1030. doi:10.1016/j.neuro.2004.03.002.

Odawara A, Matsuda N, Ishibashi Y, Suzuki, I. 2018. Toxicological evaluation of convulsant and anticonvulsant drugs in human induced pluripotent stem cell-derived cortical neuronal networks using an MEA system. Sci Rep 8:10416. doi.org/10.1038/s41598-018-28835-7

Shafer, TJ. 2019. Application of Microelectrode Array Approaches to Neurotoxicity Testing and Screening. Adv Neurobiol. 22, 275-297.

Shafer, TJ., et al., 2019. Evaluation of Chemical Effects on Network Formation in Cortical Neurons Grown on Microelectrode Arrays. Toxicol Sci. 169, 436-455.

Sharma KK, and Krishna H. 1994. Asymptotic sampling distribution of inverse coefficient of variation and its applications. IEEE Transaction Reliability. 43: 630-633.

Talcott RE. 1979. Hepatic and extrahepatic malathion carboxylesterases. Assay and localization in the rat. Toxicol. Appl. Pharmacol. 47(1); 145-150.

Thomas RS. 2012. Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol Sci. 125(1):157-174.

US EPA [United States Environmental Protection Agency]. 2020a. Agency Issue Paper: Use of New Approach Methodologies to Derive Extrapolation Factors and Evaluate Developmental Neurotoxicity for Human Health Risk Assessment. Published August 25, 2020.

US EPA [United States Environmental Protection Agency]. 2020b. Agency comments on the Exponent White Paper. Pharmacodynamic Parameters (PDPs) of Human and Rat Acetylcholinesterase (AChE) Inhibition by Direct-Acting Organophosphorus (OP) Insecticides or Active Metabolites. Published July 9, 2020.

US EPA [United States Environmental Protection Agency]. 2014. Guidance for Applying Quantitative Data to Develop Data-Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation. EPA/100/R-14/002F. September 2014.

Wambaugh JF, Hughes MF, Ring CL, MacMillan DK, Ford J, Fennell TR, Black SR, Snyder RW, Sipes NS, Wetmore BA, Westerhout J, Setzer RW, Pearce RG, Simmons JE, Thomas RS. 2018. Evaluating *in vitro-in vivo* extrapolation of toxicokinetics. Toxicol Sci. 163(1):152-169. doi: 10.1093/toxsci/kfy020.

Wetmore BA, Wambaugh JF, Allen B, Ferguson SS, Sochaski MA, Setzer RW, Houck KA, Strope CL, Cantwell K, Judson RS, LeCluyse E, Clewell HJ, Thomas RS, Andersen ME. Incorporating High-Throughput Exposure Predictions With Dosimetry-Adjusted *In Vitro* Bioactivity to Inform Chemical Toxicity Testing. Toxicol Sci. 2015 Nov;148(1):121-36. doi: 10.1093/toxsci/kfv171.

Williams ME, Wilke SA, Daggett A, Davis E, Otto S, Ravi D, Ripley B, Bushong EA, Ellisman MH, Klein G, Ghosh A. 2011. Cadherin-9 regulates synapse-specific differentiation in the developing hippocampus. Neuron. 71(4):640-55. doi: 10.1016/j.neuron.2011.06.019.

World Health Organization International Programme on Chemical Safety. 2010. WHO IPCS. Characterization and application of physiologically based pharmacokinetic models in risk assessment. IPCS harmonization project document no. 9. Geneva, Switzerland.

Yoon, M, Campbell, JL, Andersen, ME, and HJ Clewell. 2012. Quantitative *in vitro* to *in vivo* extrapolation of cell-based toxicity assay results. Crit Rev Toxicol. 42(8):633-652.