PEER CONSULTATION WORKSHOP ON THE CUMULATIVE RISK ASSESSMENT OF PHTHALATES

DECEMBER 8-9, 2010
ARLINGTON, VIRGINIA

SUMMARY REPORT

FINAL
MARCH 15, 2011

DISCLAIMER

Mention of commercial or trade names does not constitute endorsement by the U.S. Environmental Protection Agency (EPA). The views expressed in this report are those of the workshop participants and do not reflect EPA opinions or policy.

ICF International, EPA's contractor, provided logistics and note-taking at the workshop and prepared this report. This work was conducted under EPA Contract Number EP-C-09-009, Work Assignment Number 2-41, at the direction of Deborah Segal, EPA Work Assignment Manager.

CONTENTS

[This page left blank intentionally]

1. Workshop Background

Phthalate esters are a group of chemicals used in the manufacture of polyvinyl plastics and other materials including pharmaceuticals, detergents, toys, cosmetic and personal care products, medical devices and food packaging to increase flexibility and pliability. Humans are regularly exposed to phthalates in the environment, resulting in phthalates, and their metabolites, being detected in human saliva, urine, amniotic fluid, and breast milk. Growing epidemiological evidence indicates a possible association between exposure to multiple phthalates and indicators of potential effects on the male reproductive systems at exposure levels similar to background levels observed in the population.

In light of these potential cumulative exposures to multiple phthalates in the environment and their associated hazards, the U.S. Environmental Protection Agency (EPA) commissioned external experts from the National Academies of Science (NAS) National Research Council (NRC) to evaluate the issues and approaches related to performing a cumulative risk assessment of phthalates. On December 18, 2008, the NAS released their findings in a report entitled, "Phthalates and Cumulative Risk Assessment—The Tasks Ahead" (NRC, 2008). In this report, the NAS recommended that EPA perform a cumulative risk assessment of phthalates and other anti-androgens based on common adverse outcome, rather than focusing exclusively on structural similarity or on similar mechanisms of action. The NAS concluded that a sole focus on phthalates to the exclusion of other anti-androgens would be artificial and would underestimate cumulative risk.

EPA responded to the NAS recommendations by initiating the development of an Integrated Risk Information System (IRIS) Cumulative Hazard Assessment based on common adverse outcomes. The six phthalates selected include: dibutyl phthalate (DBP), di(2-ethylhexyl)phthalate (DEHP), butyl benzyl phthalate (BBP), di-isobutyl phthalate (DIBP), di-isononyl phthalate (DINP), and dipentyl phthalate (DPP). To help the EPA evaluate the recommendations made by the NAS and identify the best path forward in developing the IRIS Cumulative Hazard Assessment for Selected Phthalates, EPA convened a two-day expert workshop on December 8-9, 2010, in Arlington, Virginia. The workshop sought individual expert input, rather than group consensus, in meeting its discussion goals. The workshop agenda, a list of panelists, discussants, and the invited facilitator, and a table of workshop registrants are presented in Appendices A through C, respectively.

EPA anticipates completing an internal Agency review draft of the Cumulative Hazard Assessment for Selected Phthalates by Summer 2011. An independent expert peer review is tentatively scheduled for early Fall 2011, and the final IRIS Cumulative Hazard Assessment is anticipated to be completed in 2012. This Cumulative Hazard Assessment for Selected Phthalates will serve as a future framework for the evaluation of other compounds that cause similar adverse outcomes and is a step forward in the consideration of the risks of exposure to multiple environmental chemicals.

2. Workshop Objectives and Expectations

Dr. Kevin Teichman, the Deputy Assistant Administrator for Science in EPA's Office of Research and Development (ORD), opened up the workshop by reviewing the workshop objectives and expectations. He noted that the main goal of the workshop was to discuss and evaluate the recommendations presented in the 2008 NAS report (NAS, 2008) concerning methods for performing a cumulative hazard assessment for the six selected phthalates: DBP, DEHP, BBP, DIBP, DINP, and DPP. The specific workshop objectives and expected outputs are included in the following text box.

Workshop Objectives:

- Assess the strengths and weakness of various approaches for cumulative assessment
- Evaluate the application of approaches for existing phthalates data sets
- Discuss key considerations for extension of the identified approaches to other chemicals affecting a common adverse outcome

Expected Workshop Outputs:

- Identification of any additional published studies that might be considered in cumulative assessment
- Identification of any ongoing studies that might potentially inform cumulative assessment
- Identification of any critical data gaps that may limit application of existing approaches for cumulative assessment
- Development of a workshop report including summary of potential approaches that may be applied to cumulative hazard for phthalates

3. Day 1 – Review of Existing Phthalate Data

Presentations on the first day of the workshop focused first on setting the stage and then providing an overview of the current state of data from existing human and animal phthalate studies. Discussions following each presentation allowed the invited panelists and discussants to: (1) identify any additional published studies that might be considered in cumulative assessment, (2) identify any ongoing studies that might potentially inform cumulative assessment, and (3) identify any critical data gaps that may limit application of existing approaches for cumulative assessment.

Dr. Paul Foster of the National Institutes of Health (NIH), National Institute of Environmental Health Sciences (NIEHS) set the stage by presenting the NAS recommendations on the cumulative risk assessment of phthalates, which was followed by an overview of EPA's subsequent response, the six phthalates selected for assessment, and the various health effects collectively known as "phthalate syndrome" by Dr. Andrew Hotchkiss, EPA/ORD. Dr. Glinda Cooper, EPA/ORD, then discussed available epidemiologic studies investigating the potential associations between phthalate exposure and health effects in humans, while Dr. Earl Gray and Dr. Jason Lambert, both of EPA/ORD, focused their presentations on existing animal data for phthalates. Specifically, Dr. Gray focused on male developmental/reproductive toxicity data, while Dr. Lambert presented information on other (e.g., liver) toxicities associated with the six selected phthalates.

3.1. Overview of NAS Recommendations on Cumulative Risk Assessment of Phthalates, Dr. Paul Foster (NIH/NIEHS)

Dr. Foster, (who, along with workshop panelists Dr. Mary Fox, of Johns Hopkins University, and Dr. Chris Gennings, of Virginia Commonwealth University [VCU], had been on the 2008 NAS panel that reviewed phthalates) summarized the NAS' approach and recommendations as presented in the report "Phthalates and Cumulative Risk Assessment: The Tasks Ahead" (NAS, 2008). He outlined how the EPA commissioned the NAS, specifically the NRC's Committee on the Health Risks of Phthalates, to review the existing health effects data for phthalates, to determine whether a cumulative risk assessment could be done for this class of chemicals, and if so, to recommend potential risk assessment approaches. The EPA also asked the committee to consider the applicability of the recommendations to other chemical classes. The NAS, however, was not commissioned to conduct a risk assessment in the 2008 report.

Dr. Foster noted that the 2008 NAS committee approached this task by reviewing numerous scientific publications on cumulative risk assessment, phthalate exposure, and phthalate toxicity, as well as listening to presentations by various experts on the state of the science in the field of phthalate toxicity and cumulative risk. The committee focused on the following two questions: "Should a cumulative risk assessment be conducted for phthalates?" and, if so, "How should an assessment be conducted?" Therefore, the report was not a risk assessment of any particular phthalate or of the chemical class as a whole. Rather, the 2008 NAS committee restricted its examination to the most sensitive health effect(s) of several different phthalates, namely, development of the male reproductive system during pregnancy, as indicated by literature on animals exposed to phthalates in the laboratory. The committee recognized that not all phthalates are equipotent in their effect(s) on male reproductive system outcomes.

Additionally, the 2008 NAS committee discussed mode of action versus mechanism of action and the way in which these concepts apply to risk assessment methods. Mode of action is generally defined as a description of key events along a biologic pathway to a final health outcome, while mechanism of action is generally defined as a more detailed annotation of the biologic pathway typically at a molecular level. Cumulative risk assessment methods to date have largely dealt with chemicals with common modes or mechanisms of action. However, multiple, potentially disparate, biologic pathways can lead to a common health outcome or set of outcomes; thus, focusing on one pathway could be too narrow an approach for grouping chemicals into a cumulative assessment. As a result, the 2008 NAS committee recommended that the EPA consider multiple pathways of phthalate toxicity, and base the assessment on *common adverse outcomes* instead of a common mode or mechanism of action.

The 2008 NAS committee also assessed whether a cumulative risk assessment should be conducted. The committee determined that phthalates are present in a wide variety of products and, therefore, recommended that EPA consider multiple routes of exposure (e.g., dermal, inhalation, oral). The committee also found that according to the National Health and Nutrition Examination Survey (NHANES) conducted by the Centers for Disease Control and Prevention (CDC), simultaneous exposure to multiple phthalates in the general population, including children and adults, has been documented. Additionally, studies have shown that phthalates cross the placenta, and multiple phthalates have been measured in animal and human amniotic fluid, indicating fetal exposure.

Furthermore, the literature indicates that exposure to some phthalates causes a variety of related effects on the development of the reproductive system in male laboratory animals, particularly in rats. Known collectively as the "phthalate syndrome," these effects include, but are not limited to, infertility, decreased sperm production, cryptorchidism, hypospadias, and other related effects. The collection of effects (syndrome) is observed as one or more of these effects occurring in exposed animals, with the potential for observing different effects within the syndrome among different animals in the same

study. These effects, which are commonly caused by disruptions in androgen hormones, are similar to those included in the hypothesized testicular dysgenesis syndrome in humans. Additionally, in humans, disruptions to androgen action can lead to testicular germ cell cancer, the most common cancer observed in young men, however there is currently no experimental animal correlate for this effect (testicular tumors in rats are typically Leydig-cell-derived). Figure 1 depicts this relationship between different adverse endpoints that result from disturbance in androgen action.

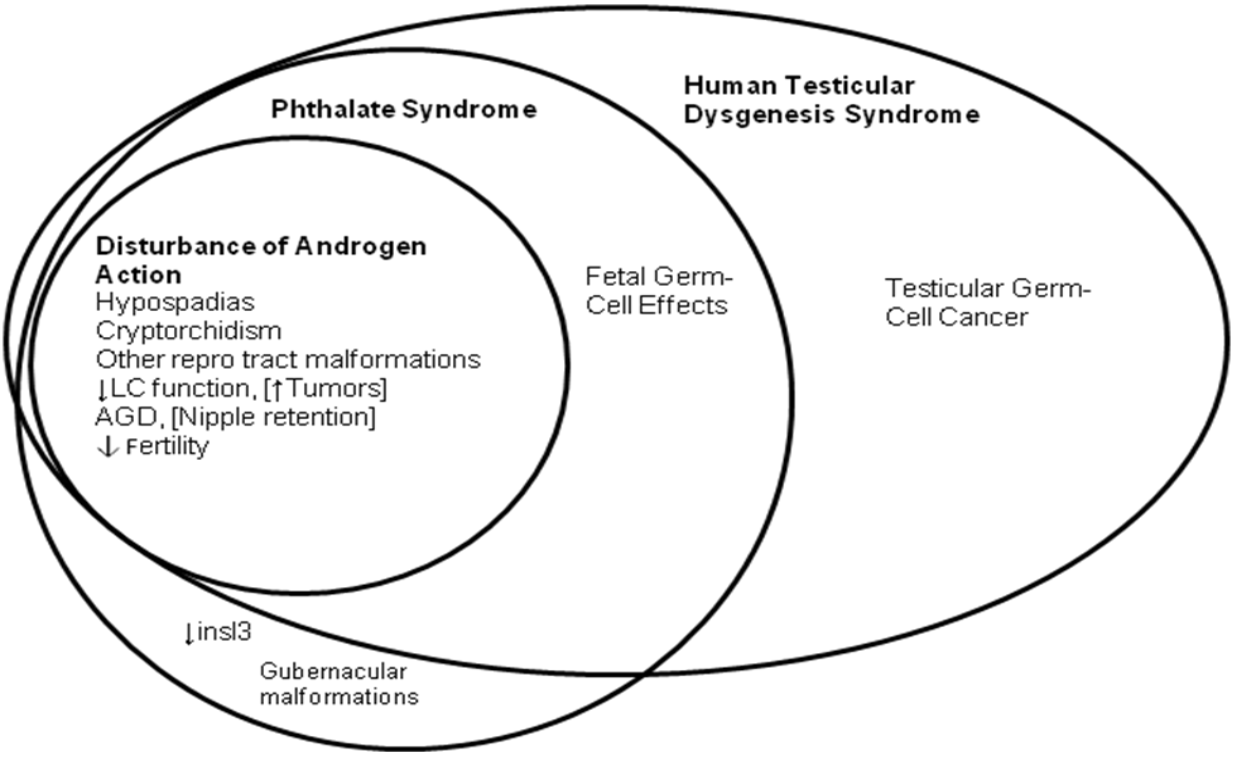

Figure 1. Relationship of phthalate syndrome in animals to other syndromes.

Dr. Foster highlighted a number of caveats in the data on phthalates. First of all, as previously noted, not all phthalates are equivalent in potency for effect in the male reproductive compartment. The most potent phthalates generally are those with straight ester side chains of 4 to 6 (possibly 3 to 7) carbon atoms; phthalates with shorter or longer side chains typically exhibit less severe or no effects. The age of the animal at the time of exposure is also critical to the severity of effects. Although all life stages can be affected by phthalate exposure, the fetus is the most sensitive, with effects generally seen at lower exposures than other life stages. Multiple studies have shown that phthalates reduce testosterone concentrations; this insufficiency is postulated to be a critical causal factor in the variety of effects observed (Figure 1), particularly if exposure occurs at times that are critical for androgen-dependent development in the male reproductive system. Testosterone concentrations, however, can be affected through different modes of action. In reproductive tissues, it is unlikely that one can differentiate an outcome due to decreased androgen synthesis versus a blocked or mutated androgen receptor. Thus, any agent that can produce either androgen insufficiency or block androgen-receptor signaling in the developing male fetus would have resulting effects that are included in the array of malformations

known to be caused by phthalates. Figure 2 depicts how various mechanisms can lead to common adverse outcomes.

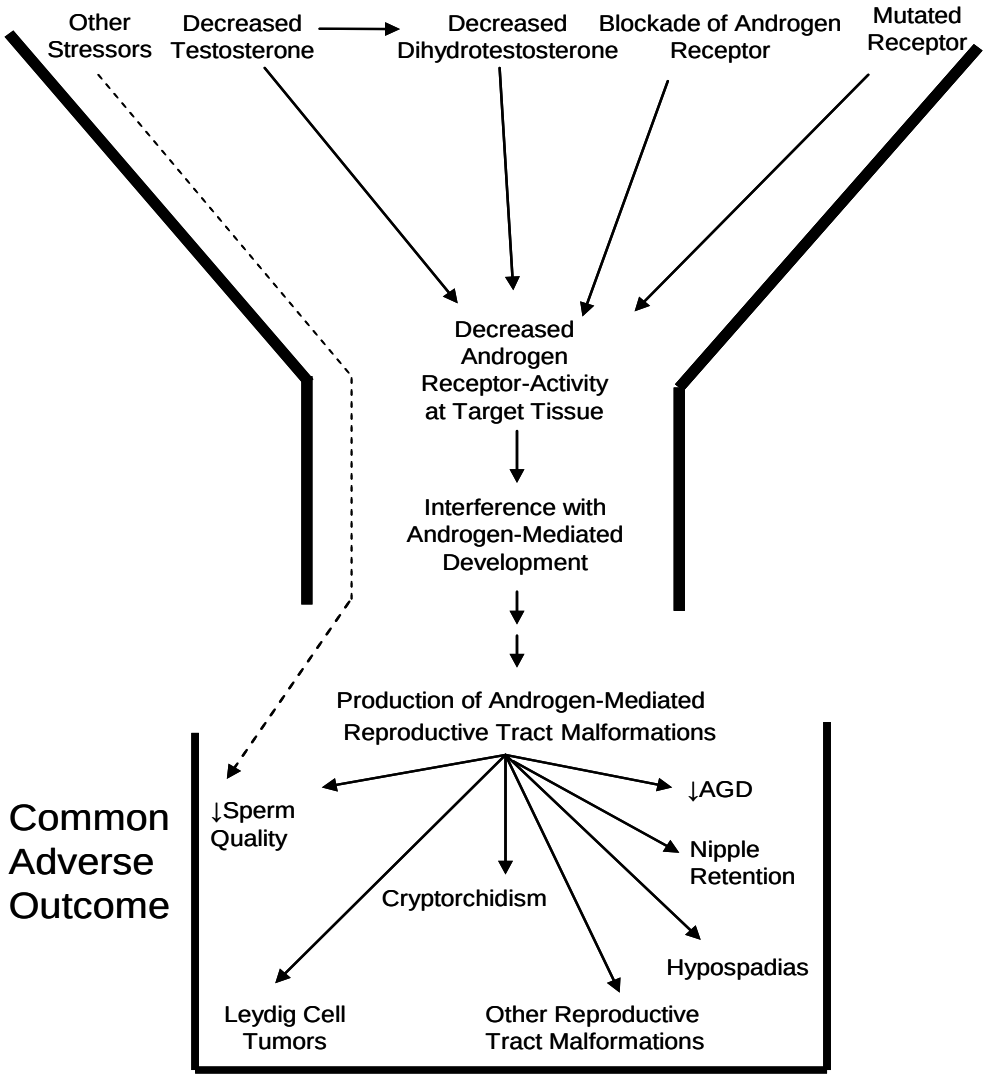

Figure 2. Various mechanisms leading to common adverse outcomes.

Dr. Foster then discussed a few of the models that describe interactions between chemicals in a mixture. He noted that if chemicals act together to produce an effect and do not enhance or diminish each other's actions, the outcome of exposure is additive, and dose addition can be applied to predict effects. Dose addition treats two chemicals as if they were dilutions of one another; that is, the same chemical with different potency, so one chemical can be added to another without changing the overall combined effect. Independent action (or response addition), on the other hand, refers to the concept that each chemical acts as if the other were not there, and causes the same effect regardless of whether the other chemical is present. (A third possibility, that of interaction between the two chemicals in a mixture, where the chemicals can enhance or diminish each other's actions, is relatively uncommon.) Historically, EPA has focused on dose addition as the default method for chemicals that affect a given

organ system. EPA mixtures guidance (EPA, 2000) has asserted, though, that if dose-addition methods are to be used, the chemicals for consideration should exhibit the same mode or mechanism of action. However, Dr. Foster noted that it can be difficult to define criteria for determining similar mechanisms of action. Additionally, EPA stipulates in its 2000 mixtures guidance document that dose-response curves of the chemicals should be parallel if dose-addition methods are to be used.

The 2008 NAS committee concluded that EPA's more recent stipulations (EPA, 2000) on when dose-addition methods should be used are too restrictive, and the differing mechanisms of action of some phthalates do not negate the appropriateness of dose addition methods. The 2008 NAS committee urged EPA to evaluate several approaches to dose addition and consider the advantages and disadvantages of each. The committee strongly recommended that EPA group chemicals that cause common adverse outcomes and not focus exclusively on structural similarity or on similar mechanisms of action. The committee also strongly encouraged EPA to consider other non-phthalate chemicals that cause androgen insufficiency or block androgen-receptor signaling in a cumulative risk assessment; a focus solely on phthalates to the exclusion of other anti-androgens would be artificial and could seriously underestimate cumulative risk. Finally, although it may appear challenging, the 2008 NAS committee was confident that sufficient data existed to proceed with the cumulative risk assessment of phthalates and other anti-androgens.

Discussion Summary

The invited panelists and discussants took this opportunity to ask Dr. Foster clarifying questions on the 2008 NAS committee's approach and subsequent recommendations to the EPA. Specifically, various panelists and discussants discussed the appropriateness of using dose addition and whether defining an "adverse" effect was possible in the case of phthalates.

Dr. Chris Borgert of Applied Pharmaceutical Toxicology, Inc. raised concern over the predictive abilities of methods focusing just on effects at common target organs rather than focusing on the mechanistic details involved with using common pathways or modes of action. He indicated that the purpose of developing more sophisticated approaches involving pathway and mode of action was to be more predictive in extrapolating to lower doses and more predictive in extrapolating beyond the species. By focusing here on common adverse effects, we seem to be losing the details useful in prediction. When chemicals are present at doses below the region producing observable effects, it is more dubious whether these chemicals will actually add up if they are not affecting the same pathway, because of adaptive or compensatory mechanisms. He stated that one cannot extrapolate to other species when one cannot compare pathways between species. Dr. Borgert asked Dr. Foster whether the 2008 NAS committee considered this issue, especially when considering the difference between dose addition and response addition.

Dr. Foster indicated that this issue was considered. The 2008 NAS committee saw these as very specific effects on development, effects about which we knew something (unlike generally classifying something as having "liver cancer" as an end effect). The committee was more concerned that if they chose a narrow biologic pathway, they would miss the opportunity to assess target tissues, which have a limited way of responding, but are impinged upon by multiple stressors. Dr. Borgert agreed with the above, but suggested the more pressing issue was related to dose addition versus response addition at very low doses. Below an observable effect range, one may be able to apply a dose additive concept if the impingement is on the same pathway, because then the adaptive or compensatory mechanisms are perhaps less likely to abrogate an effect. If you have several pathways, you need to be more in the observable range to determine which model to use, and you are also less able to extrapolate across species. Dr. Foster noted that the committee did discuss the various addition models and the possibility

that certain components in mixtures could overwhelm others. He emphasized that the committee's charge was looking at methodologies while keeping the biology as robust as possible. The 2008 NAS committee did not want to ignore agents that contribute to risk because they did not have identical modes of action, even though they affected a similar target or had a close relationship with that target. Dr. Mike DeVito of NIH/NIEHS, in response to Dr. Borgert's question, described his recent research with a mixture of 18 different polychlorinated biphenyls and dioxins, which acted by different mechanisms of action, evaluating effects on thyroid hormones. He noted that in this mixture of 18 different chemicals, at the highest dose, no chemical produced more than 5 percent of an effect, but they still showed additivity. Even after adding three additional chemicals that acted upon the same target (i.e., the thyroid), now with 21 chemicals all well below their individual NOAELs, his group was still able to predict effect at low dose levels. The challenge, Dr. DeVito wondered, is how many chemicals could be added to the model without hindering the model's ability to predict an effect through dose additivity.

Dr. Borgert also wondered if the model would stand up to many chemicals. For instance, he asked if one were to consider all byproducts of metabolism that at a high dose could be toxic to the liver and add to that all the natural toxins found in foods that could be toxic to the liver, which is hundreds of chemicals, does not the assessment come up with the inescapable conclusion that we all should be in liver failure? Dr. Borgert further noted that at some point, everything does not just add up; biology does not work that way. But the question is, where does biology cross that line? Unless the area near the observable range is the area of interest, we run the risk of deriving a model that has a rather abrupt collision with reality at some point. Dr. Borgert questioned how protective these methods would be if they fail to distinguish between theoretical hazards and real hazards. Doing experiments in the observable range does not help us determine where that line is crossed. He noted that in Dr. DeVito's experiments, 21 chemicals showed additivity, but what about 100?

Dr. Gray indicated that these are good points, but the conversation is hypothetical as to whether the model will fail at 22 or 100 chemicals. One can say we are all exposed to 100 chemicals at levels that should cause effects, but there is really no data to support that hypothesis. One cannot say we should all have liver failure without looking at the exposure levels to all the chemicals, as well as the compensation mechanisms. On the other hand, the fetus has little ability to compensate; it needs testosterone at a certain time window, and if it doesn't get it, adverse effects occur.

Dr. Borgert questioned that although the conversation was hypothetical, is it really more hypothetical than taking a no observed effect level in a rat, and applying a number of safety factors to derive a reference dose (RfD)? And then assuming that dose additivity occurs at one one-hundredth or one-one-thousandth of the RfD, and then assuming that also applies to humans? Is that not somewhat hypothetical as well?

Ms. Linda Teuschler, EPA/ORD, reminded the panel that the current phthalate assessment involves only six chemicals, and that there exist empirical data indicating that dose addition works fairly well in this instance. According to the Mixtures Guidance (EPA, 2000), one might be on shaky ground above 12 chemicals, but with 100 chemicals you certainly have a complex mixture, and in those cases, it would be more appropriate to apply whole-mixture methods. Unfortunately, the scientific community rarely has data on human exposure levels, so some amount of judgment must always be used. She noted, however, that dose addition may work for the six phthalates EPA has selected.

Dr. Kim Boekelheide of Brown University then asked Dr. Foster whether the 2008 NAS committee had discussed specifically what the "adverse" effect was in this case. Was a lowering of testosterone level, an anti-androgenic effect, if measurable, considered an adverse effect? Dr. Foster replied that fetal testosterone can be measured, and we knew that lowering of those levels at that time would result in

an adverse outcome. He noted, though, that it is not quite the same as saying lower testosterone levels is an adverse effect, but testosterone reduction could be used as a surrogate for an adverse effect. However, it is unclear how *much* testosterone would have to be decreased to see an adverse effect, as the data did not allow for an examination of points of departure.

3.2. Selected Phthalates and Phthalate Syndrome, Dr. Andrew Hotchkiss (EPA/ORD)

Dr. Hotchkiss provided an overview of the six phthalates selected for the IRIS Cumulative Hazard Assessment (i.e., DBP, DEHP, BBP, DIBP, DINP, and DPP) and the data available for each of these phthalates. He also presented an overview of the potential health effects associated with exposure to these six phthalates, including the various health effects collectively known as "phthalate syndrome."

Dr. Hotchkiss began by reviewing general information on the uses, sources of exposure, and toxicokinetics of the six phthalates. He highlighted the wide use of phthalates in consumer products (e.g., polyvinyl plastics, carpet backing, adhesives, caulks, paints, food packaging, detergents, cosmetics and personal care products), but noted that the Consumer Product Safety Improvement Act (2008) recently banned the use of DBP, DEHP, and BPP in toys. He noted that sources of phthalates exposure include, but are not limited to, emissions from manufacturing facilities, air borne emissions from phthalate-containing products, occupational exposure at manufacturing facilities, releases from municipal and hazardous waste landfills, ingestion of foods packaged in plastics, inhalation of household dust, and dermal contact with phthalate-containing products. These sources of phthalate exposure, therefore, may result in oral, dermal, and inhalation exposure to humans and could potentially occur throughout the lifetime of an individual. Dr. Hotchkiss noted that once absorbed by the body, phthalates are distributed throughout the body and may be metabolized first into monoesters and then into monoester oxidized metabolites. Phthalates are eliminated from the body through urine and feces, and the most common way of assessing human exposure to phthalates is by measuring urinary metabolites.

Dr. Hotchkiss continued that based on a review of published phthalate toxicological literature, reproductive and developmental studies dominate. He noted, however, that the literature also indicated that there are a number of other systems that may be affected by phthalates including the liver, kidney, thyroid, immune, and neurological systems. Body weight has also been affected by exposure to phthalates, and there are published studies that address the possibility that exposure to phthalates may be associated with the occurrence of testicular, liver, kidney, pancreatic, and hematological cancers. Dr. Hotchkiss further noted that the NHANES data has shown that metabolites of multiple phthalates are present in humans in the United States, with higher concentrations in children. Phthalates were also detected in amniotic fluid, indicating that they are able to cross the placenta. Epidemiologic studies in humans have also uncovered a potential link between maternal exposure and adverse pregnancy outcomes, and also adverse effects in infants.

Dr. Hotchkiss indicated that the data available for each of the six phthalates being included in the EPA's IRIS assessment varies, but across all six of these phthalates, reproductive/developmental, liver, and kidney effects were the most observed in laboratory animals. Based on the exposure-response array presented in Figure 3, reproductive/developmental effects appear to be the most sensitive endpoints. He continued that according to the literature, there are a variety of specific health effects that appear as lowest observed adverse effect levels (LOAELs) for the six phthalates, including: decreased fetal testosterone, decreased anogenital distance (AGD), increased nipple retention, increased external and internal malformations (e.g., hypospadias, cryptorchidism), decreased reproductive tissue weights, epididymal and testicular lesions, and reduced sperm production; collectively, these effects are referred to as the "phthalate syndrome," which is the spectrum of effects occurring in male rats resulting from *in*

utero exposure to phthalates during sexual differentiation. The phthalate syndrome occurs as a result of decreased fetal androgens and insulin-like 3 (insl3) hormones as a result of insl3 gene expression. Dr. Hotchkiss noted, however, that some effects for phthalate syndrome have not been measured for all of the six selected phthalates (see Figure 4). For example, hypospadias and cryptorchidism data are not available for two of the six phthalates. He also noted that study designs differed across the various published literature.

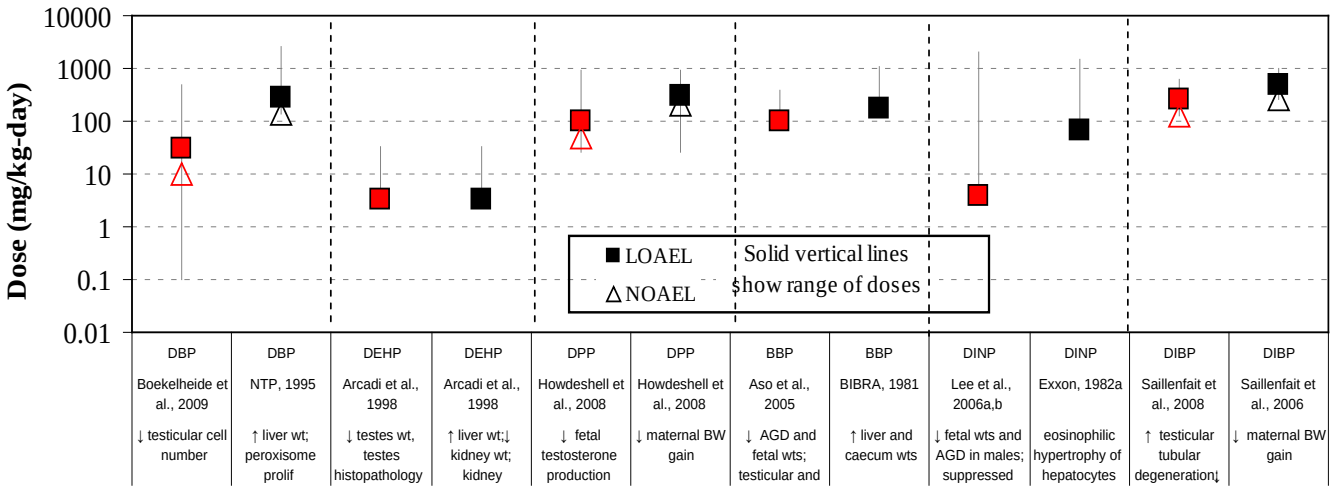

Figure 3. Exposure-response array for six phthalates. Comparison between LOAEL for developmental/reproductive endpoints (in red) and next most sensitive endpoint (in black).

	DEHP	DBP	BBP	DINP	DIBP	DPP
Hypospadias	X	X	X		X	
Cryptorchidism	X	X	X		X	
AGD	X	X	X	X	X	X
Retained Nipples	X	X	X	X	X	
Reproductive Organ Wts.	X	X	X	X	X	X
Sperm	X	X	X	X	X	X
Fetal Testosterone	X	X	X	X	X	X

Figure 4. Phthalate syndrome effects observed in published literature across the six phthalates.

Dr. Hotchkiss concluded by explaining that the phthalate syndrome observed in animals has many similarities to a hypothesized syndrome in humans—testicular dysgenesis syndrome. This syndrome

results from disruption of critical gene programming in the fetal testes. He emphasized the need to explore the possible association between testicular dysgenesis syndrome and phthalate exposure in humans further, since there is no human data available directly linking the hypothesized syndrome in humans with phthalate exposure and a number of questions remain. For example, it is unclear whether the same metabolites that are active in animals are active in humans. In addition, there are concerns that methodologies vary across the literature (limiting comparison between studies and across all six phthalates) and that urinary measurements, although the most common human exposure measurement for phthalates, have limitations and may not be an accurate measure of exposure for short-lived chemicals.

Discussion Summary

The panelists and discussants agreed with the concerns and questions that Dr. Hotchkiss presented regarding the data available and the ability to compare these data across the six phthalates. They also raised a number of other issues and/or points of clarification for EPA to consider in regards to the interpretation and comparability of the data for each of the six phthalates.

Dr. Gray raised concerns over the exposure-response array (see Figure 3) because of the differences in study design across the literature. He felt it was misleading to compare potencies across the phthalates because the studies were designed for different durations and with different intents. He also raised the point that while human exposures are generally low, data from the CDC indicate that some phthalate exposures are quite high in humans. For example, in the NHANES survey, DPP metabolites were measured at detectable levels in 29 percent of humans, although there are supposedly few if any uses of DPP.

Ms. Linda Teuschler of EPA/ORD asked if the exposure-response arrays were specific to data in the rat, and whether there were data in other species. Dr. Hotchkiss informed the panel that while there are data in other species, this particular array looked only at data in the rat.

Dr. Borgert raised the issue of whether the panel preferred, for assessment purposes, dose-response data giving the shape of the curve, or data that provide an estimated no observed adverse effect level (NOAEL). Dr. Hotchkiss indicated that the benchmark dose (BMD) approach has a number of advantages over the NOAEL/LOAEL approach. Dr. Gray noted, however, that whichever approach is used, it will influence the way animal studies are designed.

In terms of the exposure-response array for the six phthalates that presented the lowest LOAELs for developmental/reproductive effects, Dr. DeVito asked whether fetal testosterone was the lowest observed effect, followed by decreased cell number as the next most severe effect, in the spectrum of phthalate syndrome. He wondered whether there was a continuum of the dose response in this spectrum moving from low to high as there are slightly different effects as you look across studies. Dr. Gray explained that decreased Leydig cell numbers are not necessarily caused by the same mode of action or pathway as the androgen signaling pathway. It is unclear if this cellular endpoint is covered when a risk assessment is done on the fetal testosterone endpoint. Dr. Boekelheide stated that there may be separable effects on the Leydig cells versus the seminiferous cord, and this issue should be discussed initially during the course of the workshop. In some studies, including those using DBP, the dose levels producing effects in Leydig cells and the seminiferous cord were similar, but it is not certain if this is always true.

3.3. Human Data for Phthalates, Dr. Glinda Cooper (EPA, ORD)

Dr. Lambert introduced Dr. Cooper and the topic of using human data to conduct a cumulative hazard assessment of phthalates. He posed the questions included in the following text box to the invited panelists and discussants. Dr. Cooper then provided a summary of the available epidemiologic studies investigating the potential associations between phthalate exposure and health effects in humans. Keeping in mind the questions posed by Dr. Lambert, Dr. Cooper began her presentation by highlighting the strengths and weakness of using epidemiologic studies. Advantages of using epidemiologic data include the relevance of species to human health and the relevance of exposure levels to everyday and lifetime exposures in humans. Furthermore, epidemiologic data allows researchers to address human variability (such as differences in genetics and diet), which can affect susceptibility to a particular exposure. However, she noted that there are also challenges in using epidemiologic studies, which include the potential for population variability or other exposures affecting the results, rather than the results being limited to the exposure of interest. It is also difficult to separate out the effects of multiple routes or multiple exposures. It can be difficult to know the original exposure level from each respective route of concern because typically the exposure measures are integrated, as in urine samples. In other words, it can be difficult to extrapolate back to the external exposures giving rise to the measured internal levels. Furthermore, epidemiologic studies may not investigate the relevant time window. For example, one may be measuring exposure after a disease has already occurred, rather than the exposure that preceded the disease. Finally, it is difficult to study rare events without large study populations. Obtaining the consent and funding for large study populations can be challenging.

Considerations for Characterizing the Cumulative Hazard of Phthalates Using Human Data:

- Is there a possibility for elucidating potential differences in body burden by lifestage (e.g., *in utero* versus neonate versus toddler versus adult)?
- Can the data inform a relationship between maternal urinary phthalate metabolite concentrations and those in the fetal compartment?
- Are there any indications of a link between biomarkers of exposure and biomarkers of effect (suspect metabolites)?

Dr. Cooper then reviewed the existing literature on phthalate exposure levels in the human population. The majority of published studies measured metabolites of DBP or DEHP. She indicated that according to the literature, geometric mean urinary metabolite levels for various metabolites varied between 2.7 and 163 µg/g creatinine in the general population (children's levels are higher) and between 19 and 700 µg/g creatinine in occupationally exposed individuals. She also indicated that dialysis, neonatal, and other medically-treated patients with high exposure to DEHP in PVC[1] tubing may have urinary metabolite similar to those individuals that are occupationally exposed. Dr. Cooper continued that there are high correlations among related metabolites (MEHP/MEHHP/MEOHP[2] or MBP/MBzP[3]), but low correlations between those phthalates that are less related, which indicates that researchers should be able to tease out the effects of one compound compared to another. She noted, though, that it is somewhat unclear whether measures in a singular person are well correlated. In the short term, there is

[1] Polyvinyl chloride

[2] MEHP (mono-(2-ethylhexyl) phthalate), MEHHP (mono (2-ethyl-5-hydroxyhexyl) phthalate), and MEOHP (mono-(2-ethyl-5-oxohexyl) phthalate) are metabolites of DEHP.

[3] MBP (mono butyl phthalate) is a metabolite of DBP; MBzP (mono-benzyl phthalate) is a metabolite of BBP.

some evidence that correlation is fairly good. For example, in studies in which two morning urine samples were taken, the samples were well correlated (r=0.5 to 0.8 for various metabolites). In a pregnancy cohort, researchers found that metabolite levels were moderately correlated (r = ~0.5) over the 1- to 3-month study period.

Dr. Cooper went on to discuss the human evidence for the types of adverse outcomes seen in rats with phthalate syndrome, starting with the use of AGD as a measurable endpoint in humans. She discussed a study by Salizar-Martinez et al. (2004) which found that with appropriate training, AGD could be considered a reliable measurement. Swan et al. (2008) found that mothers with higher urinary metabolite levels had infant boys (around the age of 12 months) with decreased AGD. They found a 4 percent decrease in AGD per interquartile increase in the metabolites studied. Conversely, a study by Huang et al. (2009) found no association between metabolites in urine and amniotic fluid and AGD in boys. However, the girls with higher levels of MEHP and MEP had a statistically significant decrease in AGD, standardized for weight. Dr. Cooper stated that after comparing studies of average AGD in children from different countries, it is apparent that there are no population norms established for this measurement. Dr. Cooper felt that AGD is a reliable measurement in humans, but that larger population-based studies need to be completed to determine distributions. She further emphasized the need to continue investigating potential effects in girls as well as boys.

Dr. Cooper then discussed the data supporting a potential association between phthalate exposure and hypospadias (i.e., the abnormal development of the urethra where the opening develops on the underside, rather than the tip of the penis) in human boys. Studies conducted in the United Kingdom (UK) investigated phthalate exposures during three time periods (1980–1989, 1990–1996, and 1997–1998) (Vrijheid et al., 2003; Ormond et al., 2009). In Vrijheid et al. (2003), the study included 3,471 cases and looked at UK birth anomaly data. Exposure was assessed based on maternal work exposure to phthalates during the early pregnancy period. In the first time period (1980–1989), there were no associations observed between exposures and the occurrence of hypospadias. In the second period, there was a weak association between working as a hairdresser/barber and having a boy with a hypospadia, but this association was attenuated after adjusting for social class. In the third time period (1997–1998; Ormond et al., 2009), the study authors looked at UK surgical referrals (n = 471 cases, 490 population-based controls) and determined that maternal work exposures to hair spray and phthalates, and work as a hair dresser, were each associated with an increase in hypospadias, after adjusting for income, maternal smoking, birth weight, and folate use.

Dr. Cooper also summarized one human study investigating the association between cryptorchidism (i.e., undescended testes) and phthalate exposure. She explained that cryptorchidism can occur at birth or by 1 year of age, but can also be acquired during childhood. In a pregnancy cohort study from Denmark and Finland (nested case-control, case-cohort design), researchers assessed phthalates exposure using breast milk samples collected from 1 to 3 months of age. There was no association between levels of phthalates in breast milk and incidence of cryptorchidism. However, authors found that lower levels of free testosterone in the infants were associated with higher levels of MBP and MEP[4] metabolites in the breast milk.

As evidenced, developmental data are limited. Dr. Cooper noted that there are no known studies of hypospadias using urinary measures of phthalate metabolites and there are no studies of acquired cryptorchidism in relation to potential endocrine disruptors. However, Dr. Cooper concluded that the data that does exist, particularly related to testosterone effects, are concerning.

[4] MEP (monoethyl phthalate) is a metabolite of DEP.

Dr. Cooper went on to summarize existing epidemiologic data regarding adult men and exposure to phthalates. In a Swedish study of military recruits aged 18 to 21 (Jönsson et al., 2005), authors measured urinary phthalates as well as sperm parameters and sperm DNA, testosterone, estradiol, luteinizing hormone (LH), follicle-stimulating hormone, and inhibin. There was little association between urinary metabolite levels and any parameter measured, although increased levels of MEP were slightly associated with a decrease in LH and a decrease in sperm motility. However, Dr. Cooper explained that this was a young, healthy population, so it is not surprising that there was little effect seen in this setting.

A series of studies conducted with infertility patients at the Boston General Hospital in the United States measured similar outcomes to the Swedish military study. Authors found that low sperm concentration was associated with the highest quartile of metabolite levels (in one study, OR = 3.3; Hauser et al., 2006). There were also increases in sperm DNA damage with increasing levels of certain metabolites; this was most pronounced for MEHP. Increased levels of MEHP were also associated with decreased levels of testosterone and increased levels of free testosterone (Meeker et al., 2009). In another study by Pan et al. (2006), exposure levels were higher than in the general population. Authors reported an inverse correlation between MEHP and MBP and testosterone levels. They could not differentiate the effects of metabolites from each other since they were correlated.

Based on these studies and the robustness of the data, Dr. Cooper stated that she had considerable concern regarding the health effects observed in adult men as a result of exposure to phthalates.

In the final portion of her presentation, Dr. Cooper discussed various other effects seen in epidemiologic studies of phthalate exposure. She indicated that there have been studies investigating phthalate exposure and precocious puberty (i.e., the development of breasts before the age of two or puberty before age eight). However, two studies found little association between this condition and measured urinary phthalate metabolite levels. A study investigating the timing of puberty (Wolff, 2010) is in its early stages, but has found later signs of development with some metabolites and earlier development with others after one year of follow up. Meeker et al. (2009) have also reported a relatively strong association between phthalate exposure and risk of preterm births. Other reports have suggested that obesity and diabetes, immune-mediated conditions such as asthma, and neurodevelopmental problems and disorders such as attention deficit disorder may be associated with phthalate exposure.

Dr. Cooper concluded her presentation by stating that based on the human data, there is reason to think beyond effects seen in infants and reason to think about other metabolites (e.g., MEP from DEP). There are also other outcomes besides fetal/infant development that could have broad public health significance. She identified a number of issues that must also be addressed in order to build confidence in using human data to characterize the cumulative hazard of phthalates, including evaluating the interpretation of urinary metabolite measures, weighing relative potency effects among metabolites, evaluating the sensitivity of effects for epidemiology studies (e.g., lowered testosterone versus other, rarer clinical conditions such as cryptorchidism which would require a much larger study) and their relevance to risk assessment. She also called for including broader exposures than just phthalates.

Discussion Summary

The invited panelists and discussants discussed a variety of topics during this discussion session as described in the following sections.

Study Design

The expert panelists and discussants asked Dr. Cooper several questions regarding the study designs of the various epidemiologic studies she described during her presentations.

Dr. Rochelle Tyl of RTI International raised the question of how to deal with a study population in which only 14 percent of males donate their sperm, as was the case in the Swedish study by Jönsson et al. (2005), and she wondered whether this was a self-selecting population, where anyone who had the least concern about their health would not participate. Dr. Cooper explained that this issue might be a problem in many studies, but in this case one could see variability in exposure and a relationship between exposure and semen characteristics within the people who donated. She commented that as a population, 21-year-olds likely have not thought about or are not concerned about their semen characteristics; therefore, the low participation rate is unlikely related to knowledge of the exposure or knowledge of the outcome.

Dr. Tyl also asked about the variability in AGD among countries for females, as shown in one of Dr. Cooper's slides. Dr. Cooper stated that based on the available literature, it is unclear whether there is actual variation between countries, since the study samples were very small (32 girls).

The panel also raised questions about whether authors of the studies on AGD corrected for body weight, body length, and gestational age. Dr. Cooper indicated that authors used adjustment factors for body weight or length. Dr. Boekelheide asked if for cryptorchidism, the gestational age was considered, since premature births have more problems with cryptorchidism. Dr. Cooper replied that she thought all of the studies involved term births and had standardized protocols. Dr. Boekelheide also noted that there have been more pre-term births in the last few decades and that historical studies may not have grappled with the problem.

Dr. Boekelheide also stated, as a point for future discussion, that he had always used the signal for an MEP effect as a false positive, related to the weakness of the data set, but it appeared to him that Dr. Cooper was taking a different approach with MEP effects.

Dr. Borgert asked that if effects are a result of the large number of chemicals to which an organism is exposed, what can be said about a study of a single chemical without controlling for the myriad of other chemicals to which people are exposed that may produce similar effects? He questioned how one could make conclusions about individual chemicals without controlling either for other chemicals or for cumulative exposures. Dr. Cooper responded that unless the other chemicals are highly correlated to the exposure of interest, there should not be confounding. One could go back to the authors and see what data they have, since some of these studies looked at other endocrine disruptors along with phthalates.

Dr. Gray spoke about inconsistencies in endpoint measurement. He noted that researchers are measuring AGD in different ways and hypospadias are often scored differently depending upon the physician. As a result, it is unclear whether hypospadias incidence has been increasing or decreasing. These inconsistencies in measurement pose a challenge for epidemiologic studies on these endpoints.

Sampling Time

The expert panelists and discussants also discussed potential issues with sampling time and the ability to appropriately characterize an exposure. For example, Dr. Gray raised the issue of the meaningfulness of a single biomonitoring sample for chemicals like phthalates that have very short half lives. Since internal phthalate levels can change by an order of magnitude within a given day as well as from day-to-day, perhaps a 24-hour urine sample would be a composite measure that could give us better information.

Dr. Paul Lioy of the University of Medicine and Dentistry of New Jersey followed up by adding that Dr. Matt Lorber of the EPA has some good data on the differences in phthalate metabolite concentration based on the time of day (estimated from dietary exposure). Dr. Lioy noted that point measurements of biomarkers might not be the right method for establishing exposure in a population. He suggested that

researchers do multiple measurements, and thought that another workshop similar to this one would be essential in designing an epidemiologic study which clearly articulates the exposure. Dr. Lioy stated that the variability in exposure and exposure time are just as important as variability in population. Dr. Cooper responded by reminding the panel that the strongest study on infant testosterone levels had multiple breast milk samples collected from 1 to 3 months, making this a more integrated exposure measurement.

Exposure

Dr. Lioy also commented on the importance of determining which outcomes are important, and then linking exposure measurements to those outcomes. He noted that if one is examining effects in a fetus, one must be sure they are looking at exposure for the correct time period. Also, he suggested researchers may need to measure levels of exposure before a mother gets pregnant and/or measure the breast milk for a longer period after birth. Dr. Cooper again emphasized that there are a number of maternal-child cohort studies around the world (e.g., the National Children's Study) that are planning to address these kinds of questions.

The panel then discussed the research surrounding phthalate exposure and obesity. Dr. DeVito noted that if phthalate exposure is primarily dietary, it is difficult to make conclusions about obesity based on the NHANES data. If a person eats more food, that person is increasing their exposure to phthalates, but is also consuming more calories. For this reason, blood levels of phthalates may be correlated, but may not be accurate predictors of obesity. Dr. Cooper pointed out that the metabolites that are being associated with obesity are very specific. It would be a clearer signal that it was dietary if researchers were seeing associations with all phthalates. Dr Lioy suggested that this might be a result of a diet that has certain phthalates regularly and others periodically. Furthermore, there is a lot of variability in diet among people, which will affect exposure. Women who are pregnant alter their diet throughout their pregnancy. Any simple statement about dietary exposure based on a biomarker being there or not being there needs a lot more work before being accepted. Dr. Tyl suggested that integrated sampling over time may help account for variability in diet.

Dr. Cynthia Rider of NIH/NIEHS inquired whether it is less likely to have confidence in the phthalate levels in spot samples due to their variability. She wondered if one could have confidence that the observed connection was actually there, or whether one may have caught a peak in the levels. Dr. Cooper responded by saying that in general, yes, researchers worry about exposure misclassification. However, when one has exposure misclassification that is not related to the outcome, attenuation of the observed effect is expected, which would mean the true relationship was larger than what was estimated.

Dr. Foster asked the panel to address the issue of human exposure measurements. He asked who should be protected; for example, should the population with median or mean exposure or those in the 95^{th} percentile of exposure be protected?

Dr. Gray stated that since phthalate distributions are skewed, the mean or median is quite low, and one should focus on the 95^{th} or 97^{th} percentiles as these people are exposed to levels four to five orders of magnitude higher and are of immediate concern. Unlike ecotoxicology, we are not concerned whether a population can sustain itself; here we are worried about all the individuals. Dr. Lioy stated that some of the original EPA work focused on the 90^{th} percentile and above, which is a more traditional approach for many types of risk assessment. He felt that this population would have more variability in it, whereas people falling above the 95^{th} percentile might have atypical behaviors. For example, some women spend much more time going to the hairdresser than others.

Identifying Sensitive Subpopulations

Dr. Foster then raised the question of whether researchers have enough data to identify pregnant women and their fetuses as the exposed population of interest.

Dr. Lioy said that he was using the existing fetal toxicology as the basis of his comments and that if more data on endpoints beyond those in the fetus surfaced, his opinion would change. If one focused on the pregnant woman, one could ask, what would be the most important products contributing to risk? This would then target the population at highest risk and help define the types of epidemiologic studies needed.

Dr. Boekelheide stated that women in their first and second trimester were the population of concern in terms of fetal reproductive outcomes, since peak testes production occurs at 16 weeks gestation. However, the neonatal intensive care population is also very vulnerable, and one should not forget those that are occupationally exposed.

Issues with Using Human Data for Conducting Cumulative Hazard Assessment

In addition to specific questions about study design, sampling time, setting health protection levels, and identifying sensitive subpopulations, the expert panelists and discussants also discussed several issues and challenges for EPA with conducting a cumulative hazard assessment using human data from epidemiologic studies.

Ms. Teuschler commented that she was not sure if any of the studies presented generated an exposure-response curve; however, she noted that if they did, one could do BMD modeling. Parent compound data would be needed, however, which can be challenging to obtain when extrapolating back from the metabolites. If an exposure-response curve cannot be generated from the data, a NOAEL or LOAEL would be helpful. In addition, she noted that in cumulative risk assessment, the EPA is looking at multiple routes and pathways of exposure. She noted that typically the National Institute for Occupational Safety and Health is concerned with the occupational exposure levels, while the EPA is concerned with the environmental exposure levels. However, if EPA is doing cumulative assessment, it is unclear where the boundaries are for dietary, occupational, environmental, or water exposure levels, and it is unclear how EPA will handle all of that.

Expanding upon Ms. Teuschler's comments, Dr. DeVito explained that while typically a urinary marker of metabolites is measured, EPA regulates on intake. The challenge is going from urine metabolites to actual exposure using pharmacokinetics (PK). Dr. DeVito said that he had been looking for PK models, but could only find one for DEHP. Since the literature is not filled with phthalate PK models, he pointed out that it would be hard to use human data in a quantitative way in the year allotted to finish the phthalate IRIS assessment. Dr. Cooper, however, was under the impression that there are existing methods to extrapolate from urine metabolites to the parent compound and that there was agreement in these measures.

Dr. Lioy cautioned about methods for back extrapolating exposure, emphasizing that it is hard to go from the body to the route of entry. There may be methods available, but it all depends on which route of exposure was first. He posed the question of how you match phthalate exposure up with spot sample biomarker data when there are phthalates with short half lives. He noted that the uncertainty with the back extrapolation approach can be extremely large.

Dr. Foster asked how maternal urine was reflective of fetal internal dose. Dr. Gray responded by saying that in his studies, they did not have maternal urine; rather, Dr. Gray noted that they had dose-response data and data from the amniotic fluid in rats exposed to DBP and DEHP, and compared that with amniotic fluid data in humans. Based on this data, the highest levels of metabolites in the human fluid

for MBP and MEHP were different by an order of magnitude or so. There were no urine samples. Dr. Gray indicated that the CDC is now looking at phthalate metabolites in blood, but it is difficult to find anyone who can measure tissue levels of phthalate metabolites well.

Dr. Barry McIntyre of NIH/NIEHS inquired whether the panelists had a feel on how differential metabolism of phthalates may affect urinary metabolite levels. Dr. DeVito said he had not seen a lot of PK data on these levels and that there are a lot fewer metabolite identification data in rodent urine in the literature, so it makes it challenging to answer that. Dr. Gray mentioned that in the collaborative studies he is doing with the CDC, there may be sufficient data to model metabolism. With some chemicals, it is suspected there may be no detectable metabolites in the urine. Dr. Gray also stated that he would not assume that what one finds in urine measurements would be the same as what one would find in blood or tissues.

Dr. Foster then questioned whether having a higher level of a metabolite, such as MEP that has been shown in studies to be relatively inactive, might affect the kinetics of the other metabolites. Dr. Boekelheide stated that no one knows the answer to that, and that one must keep in mind that the route of exposure can be different for MBP and MEP (e.g., dermal absorption rather than oral). Dr. DeVito asked if anyone knew about human exposure to the phthalate metabolites themselves; this could potentially disturb reverse pharmacokinetic calculation methods.

Dr. Borgert asked whether the human is informative given all the variability we have heard discussed that would reduce the significance of any associations. Specifically, he asked whether measurements of metabolites from single chemicals from a certain class yield any meaningful information when we are really trying to look at cumulative hazard. Dr. Gray commented that if you take a single sample of six phthalates and you know the level varies throughout the day, at some point the level will be higher than you measured. One would not know what the peak height was or duration or area under the curve, so one would be underestimating the levels. Dr. Borgert suggested that looking at the data for one phthalate when humans are exposed to many is rather like doing a study of alcohol and focusing on one brand of beer, without taking anything else into consideration. Dr. Foster noted, however, that there are not hundreds of chemicals causing these specific effects.

Additional Studies and Research to Consider

Dr. Lioy noted that the data presented by Dr. Cooper clearly indicate that humans are exposed to phthalates. He suggested engaging in new studies with testable hypotheses. Researchers know which specific endpoints to examine for phthalates and although there are other confounders, if research is tailored to the correct endpoints, one could create meaningful exposure assessments and epidemiologic studies.

Dr. Rider further suggested that examining total phthalate doses and corresponding effects in animals might be helpful. She added that many other chemicals with similar effects are not as ubiquitous as the chemicals of interest. Dr. Cooper followed up by saying that there is a group in New York that combines low and high molecular weight phthalates. She recommended investigating the advantages and disadvantages of the total phthalate approach. Dr. Boekelheide commented that there is a concern in terms of the rules of exclusion for phthalates in a total phthalate study, since one researcher created a phthalate index, but included MEP, which is not active in the rat model.

Ms. Teuschler indicated she was uncomfortable with the idea that the only thing the panel could draw from the epidemiologic data was that humans are exposed to phthalates. She commented that a lot of the epidemiologic studies have reference populations that control for other background exposures. In addition, Dr. Cooper's presentation was only a half hour, so it would be wise to do a more careful evaluation of the data before one says there were no effects seen. Dr. Gennings expressed similar

concerns and urged panel members to be wary of concluding that chemicals have no effect. The sample size of the studies may have been small, and it is hard to say if power was addressed in these epidemiologic studies. It is important to have a large sample size to see effects. She also noted that it would be valuable to look at other effects besides the phthalate syndrome.

Dr. Foster then asked the panel to discuss how strong they thought the effect data were, or whether data were only qualitative. Dr. Gray began by discussing AGD. He noted that there were no consistencies among studies, but that the decreased AGD in girls indicates there is an effect, but it makes no biological sense. He indicated that one cannot come to a conclusion here. The levels of exposure, however, are lower than the doses given to lab rats, and AGD does not appear to be the most sensitive endpoint in the rat. Dr. McIntyre then reminded the panel that since small fetuses tend to have smaller AGD, researchers will somehow need to control for that in human studies, which are much more heterogeneous.

Later in the discussion, Dr. DeVito questioned whether AGD is a relevant endpoint in humans and whether the 4 percent change per quartile in AGD is anything to be worried about. Dr. Boekelheide thought the AGD effect was found in association with MEP, which is not toxicologically active in animals, which causes uncertainty about the AGD effect. Dr. Cooper explained that in her opinion, there did not seem to be a lot of interest in measuring and using AGD in the human population.

Dr. Boekelheide discussed the species specificity of the reproductive effects seen. He noted that mice are more resistant to anti-androgen effects of phthalates than rats are; however, there are consistent effects in some of the fetal testes outcomes. There are at least two groups looking at these effects. Although his own work is not published yet, there is a consistent seminiferous cord effect, but no anti-androgen effect in human fetal testes. Dr. Boekelheide also mentioned that measuring AGD in adult men is very difficult. Given that relative AGD continues in rats through adulthood, we tested groups of fertile and infertile men, about 100 in each group. When this was attempted, the data were so noisy (interpersonal variability) that it drowned out any potential association with fertility or infertility. He indicated that this measurement is really constrained by body weight, and any association is just not there.

Dr. Gray followed up by saying that studies in the literature have reported testicular effects in mice exposed to some of the phthalates at higher doses during pregnancy, and that some mouse strains were sensitive to DEHP as pubertal animals. Dr. Gray also mentioned that a new mouse study measuring AGD would find effects in mice at the higher doses. Dr. Gray explained that in marmoset studies (with five animals), there were subtle effects, but no hypospadias or testosterone effects from phthalate exposure.

Dr. Foster then asked the panel to discuss the potential genotoxicity of phthalates. Dr. DeVito noted that DEHP is negative in genotoxicity tests, but continued on to say that it would be necessary to take a closer look at the papers before dismissing genotoxic effects. Dr. Boekelheide further added that there are potential targets of phthalates in humans that do not exist in rats, and agreed that the idea of genotoxicity could not be thrown out.

3.4. Animal Data (Including Mixtures Data) for Phthalates: Male Developmental/ Reproductive Toxicity, Dr. Earl Gray (EPA/ORD)

Dr. Hotchkiss introduced Dr. Gray and the topic of how animal data (including mixtures data) for phthalates, specifically looking at male developmental/reproductive toxicity, could be used for conducting a cumulative hazard assessment of phthalates by posing the questions in the adjacent text box. Dr. Gray then opened up his presentation by reviewing how ubiquitous phthalates are and noting that if researchers only focus on the effects of phthalates on androgens and androgen signaling, then researchers will be excluding other reproductive effects that occur by different modes of action.

Should Cumulative Hazard Using Animal Data be Characterized:

- Based on phthalate syndrome as a whole as the critical effect?
- Based on the most sensitive outcome in phthalate syndrome 'shared' among individual phthalates as the critical effect?
- Based on mixtures data?
- Based on individual phthalates data?

He then reviewed how exposure to phthalates results in effects in a number of mammalian species, but noted that there are very few *in utero* data available in species other than the rat. He described the taxonomy of the animal species sensitive to phthalates and the effects observed in these species. Specifically, he noted that rats are considered more sensitive to phthalate exposure than mice. Hamsters can also be affected, but less so than rats or mice. Additionally, while there are some studies in primates showing significant effects, these effects were considered to be not statistically significant because of natural variability in that species.

Dr. Gray continued on to say that not all phthalates are reproductive toxicants. The structure-activity relationship (SAR) varies with the chemical structure of the phthalate. For example the SAR varies with length of the ester side chains, position of the two ester side chains, and other characteristics. The SAR for reproductive/developmental toxicity appears to be similar for fetal and pubertal male rats, i.e., the phthalates that seem to be potent in the fetal rat appear to be potent in the pubertal rat. The SAR does not hold for liver effects, and it is not known if this SAR predicts effects on the ovary during pregnancy, or the female "phthalate syndrome", which occurs at higher doses than in males.

He noted that phthalates reduce androgen-like and insl3 hormones, which leads to subtle demasculinization of the male reproductive tract at low doses and severe malformations in hormone-dependent tissues. It is unclear exactly how this occurs, which contributes to the uncertainty in interspecies extrapolation. It is known, however, that phthalates reduce fetal testis expression of genes involved with steroid transport, steroid synthesis, and insl3 peptide hormone synthesis. Dr. Gray displayed and explained scans of Leydig cells from male rats exposed to DEHP, DBP, or BBP. At gestation day 20, there were large clusters of these cells, which were immature and appeared to be producing less testosterone than controls. Leydig cells also produce insl3 hormone, and exposed animals have no gubernacular ligament or elongated ligaments, causing altered locations of the testes compared to controls. The gubernacular effect is unique to phthalate syndrome, and is not seen with exposure to androgen receptor antagonists or other chemicals.

Dr. Gray then explained the process of sex differentiation in fetal development of animals. Prior to sexual differentiation, all mammalian fetuses have the potential to develop as either male or female having both duct systems. Normally, in the male fetus, the testis produces hormones, which cause the

male tract to develop and the female tract to regress. Conversely, in the female, the absence of these hormones causes the male tract to regress and the female tract to develop. There is then a period *in utero* when male hormones can be disrupted, which produces phenotypic females. If hormones are not disrupted, the animal continues developing as a male and the testes produce testosterone. In the absence of or alteration of testosterone, insl3, or other hormone production, therefore, there can be profound phenotypic changes in mammalian species. Dr. Gray concluded that this pathway, and possibly phthalate effects, is highly relative to humans because this pathway is conserved in all animals and critical for human reproductive development.

Dr. Gray highlighted selected congenital abnormalities of human sexual differentiation (e.g., abnormalities of testosterone synthesis) in humans that result in recognizable phenotypes, including those syndromes that most closely resemble what is going on in the fetal phthalate syndrome. For example, the androgen insensitivity syndrome causes a person to look and behave like a normal female, although the person has testes and XY chromosomes. There are also timing defects in which all the appropriate organs and hormones are there, but the timing is late compared to the critical period resulting in abnormal differentiation and development of ambiguous genitalia in men. He reiterated the importance of this pathway in humans.

Dr. Gray then posed the question of how to screen and test phthalates for risk assessment. He questioned whether it was necessary to conduct multigenerational studies using thousands of animals to determine effects resulting from phthalate exposure, or whether there is a method available to screen phthalates for risk assessment using *in vitro* techniques or short-term *in vivo* assays. He indicated that *in vitro* assays and high-throughput screening tests of phthalates have produced negative results. These assays do not predict what can happen *in vivo* and can provide false positives or irrelevant signals. At the current time, therefore, *in vitro* assays cannot be used to screen phthalates for risk assessment. Short-term *in vivo* assays with fetal or pubertal male rats are essential for classes of chemicals like the phthalates. These "fetal phthalate screens" (FPS) use fewer animals and can make reasonable predictions about the adverse consequences of *in utero* exposure to phthalates. Dr Gray's group hypothesized that these assays can be used to predict that phthalate syndrome seen in F1 rats. Fetal testosterone levels should be appropriate as a quantitative endpoint, as fetal androgen disruptions are causally linked to the downstream tissue effects and malformations. Researchers do not know the primary initiating event, so there is some uncertainty in extrapolating effects observed in animals to humans. However, he stated that scientists are rarely 100 percent certain about the relevance of effects to humans for any chemical.

FPS assays expose a small number of pregnant females to a single high dose of phthalates for five days during the critical period of sexual differentiation. Results from the FPS assays were sorted into positive and negative, with positive results evaluated in follow-up dose-response studies. Key phthalates that are thought to be important are followed up with a postnatal study (this is being done for DPP) to determine if reduction in testosterone production and gene expression are predictive of the severity of phthalate syndrome in male rat offspring. The hypothesis is that phthalates cause lesions in male rats at certain doses. Dr. Gray then summarized the results of the phthalates that have been tested in FPS, but noted that the results are ongoing and unpublished.

Dr. Gray continued with a summary of data (Howdeshell et al., 2008) that showed fetal testosterone effects of various phthalates. DEP did not appear to affect testosterone; DEHP, DIBP, DINP, BBP, and DBP were roughly equivalent in potency; and DPP was three times more potent. Howdeshell et al. (2008) also measured the ED_{50}s (the dose required to produce a specific effect in half of a test sample) of these chemicals and used that information to predict how they would behave in a mixture study. Dr.

Gray's group proposes that one could use these data for risk assessment and would not necessarily need the postnatal data.

Dr. Gray's group has also been measuring some of the genes associated with steriodogenesis and insl3 peptide production one gene at a time. They have since stopped doing it that way and are using polymerase chain reaction arrays, incorporating data from studies that are currently unpublished. Following this step, they will have a way to determine which genes are affected by which phthalates and will have a way to compare the dose-response data of these genes to the testosterone production. So far, the data indicate that the dose-response data of testosterone provide a lower ED_{50} or BMD than do the genes or the other endpoints, indicating this endpoint is the most sensitive.

Dr. Gray continued that the male phthalate syndrome in rats results in a number of effects including but not limited to agenesis, undescended testes, testicular lesions, and hypoplasia of the epididymis. A litter of animals can all have different effects but still have the phthalate syndrome; one should classify by syndrome rather than the specific effect. Unfortunately, in most papers you cannot tell from data how effects are distributed within litters.

Dr. Gray then discussed effects of AGD from phthalate exposure. To the best of his knowledge, risk assessors have never used AGD in a risk assessment. Dr. Gray explained that based on data from Hotchkiss et al. (2004), once the AGD is affected, it will remain affected during the duration of an individual's life. In addition, the shorter the AGD, the more likely a male animal is to have other effects such hypospadias or a vaginal pouch. Barlow et al. (2004) published a study looking at the effect of DBP exposure through development of animals. They did not find a point of departure (POD) because some animals with normal AGDs had malformations, so AGD is not the most sensitive endpoint for phthalates.

Furthermore, some phthalate studies showed hemorrhagic testes in male rats, which is a unique effect. Another anti-androgen effect involves the retention of female-like nipples in males. However, this measure is subjective and not standardized across laboratories. It would be difficult to defend a risk assessment based on this data alone, but fortunately, body weight does not confound this measure and this measure is predictive of what other results that may be observed down the line. For example, the number of nipples in infant male rats is associated with severe malformations seen in adult necropsy.

Other effects observed include those described in studies of DINP which resulted in malformed testis and epididymides in exposed rats. The study determined that testes lesions can occur from the phthalates direct effect on them, or from the loss of the epididymis, which leads to fluid pressure atrophy. The testes appear normal until after puberty. Even though this occurs at a high dose, it is clear that DINP disrupts sexual differentiation.

When looking at the dose-response data of all the endpoints in the phthalate syndrome at different doses of DEHP, it appears that fetal testosterone production is a useful endpoint (see Figure 5). If you take the ED_{50} from any one of the other effects, it is always higher than that of the testosterone endpoint or as the phthalate cumulative response.

Dr. Gray then presented dose-response data for some of the individual chemicals. For example, a recent study from the National Toxicology Program (NTP; Blystone et al., 2010) identified a lower NOAEL (by an order of magnitude) for gross testicular and epididymal alterations than previous studies by retaining extra animals (litter mates) and evaluating them for gross lesions instead of discarding them.

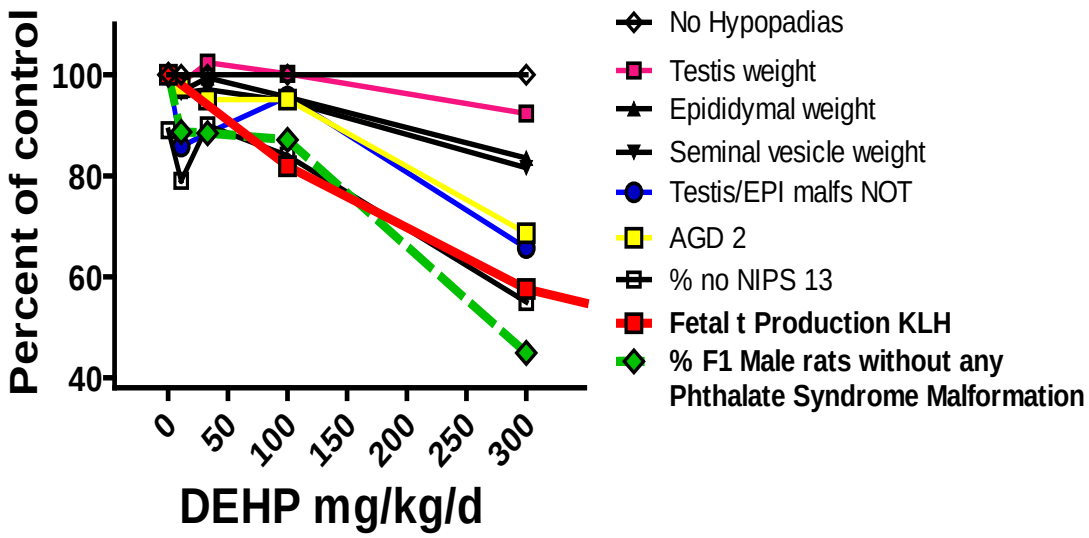

Figure 5. Critical "adverse effects": Percent of F1 male rats with the phthalate syndrome (Gray et al, 2009) and fetal androgen production (Howdeshell et al., 2008).

Dr. Gray continued with a discussion of effects of phthalate exposure on female animals. He indicated that they have seen females with uterus unicornis or complete uterine agenesis with occasional vaginal agenesis. This effect resembles a relatively common human syndrome called the MRKH (Mayer Rokitansky Kuster Hauser) syndrome. The animal model may be a useful animal model for studying MRKH, as we do not know about the genetics or mode of action of this syndrome. The critical exposure period in males is gestation day 16 through 19, but it is unclear if this is the critical period for females.

In a study with five phthalates, there was an unexpectedly high incidence of reproductive tract malformations of female rat offspring. The upper vagina was formed, but the lower vagina and a uterine horn were missing. This is not necessarily the result of an androgen-testosterone relationship, and it may be occurring much earlier in development.

Dr. Gray went on to discuss the studies his laboratory has conducted on binary mixtures and mixtures of five phthalates. The objectives of the mixtures research were to determine how the chemicals with similar or dissimilar mechanisms of toxicity interacted during sexual differentiation and to provide a framework for deciding what chemicals to include in a cumulative risk assessment. The working hypothesis is that chemicals that disrupt the development of a common reproductive tissue or system during sex differentiation will produce dose additive responses regardless of the molecular mechanism or signaling pathway that is disrupted.

The simple binary studies were not meant to produce dose-response data. Mixture studies with two phthalates (e.g., BBP and DBP, which have a common metabolite) were completed in order to discern whether the chemicals follow dose or response addition. In a binary study on other anti-androgens, the androgen receptor antagonists vinclozolin and procymidone produced dramatic effects (similar to phthalates) when combined.

In a five-chemical mixture study, DBP, DIBP, BBP, DEHP, and DPP were administered as a mixture on gestation days 8–18 and fetal testosterone was measured. The testosterone reduction was well predicted with the dose addition model, but not the response addition. Dr. Gray concluded by saying that dose addition should be the default model because it is supported by existing data, response addition does not explain the results for many of the effects seen in the studies, and it is consistent with

biology of hormone action. All these chemicals are disrupting a common pathway, and what the tissue sees is a reduction in androgen receptor or bound androgen, and in both cases the androgen-dependent gene is attenuated. What is happening at the level of the gene and the target tissue is absolutely identical whether it is a phthalate or an androgen receptor antagonist. The androgen receptor antagonists do this by preventing testosterone from binding, so there is less androgen-activated gene, and testosterone level inhibitors like the phthalates do this by reducing testosterone levels. The developing tissue cannot tell the difference.

Discussion Summary

The invited panelists and discussants had several questions for Dr. Gray, as well as used this opportunity to continue the discussion on dose addition and using human data to characterize the cumulative hazard of phthalates. The group also addressed the questions posed to them by Dr. Lambert at the beginning of this presentation.

Dr. McIntyre began the discussion by asking Dr. Gray if his lab had performed histopathological examination of animals exposed to phthalates and wondered whether these results might be more sensitive than the inhibition of testosterone. Dr. Gray replied that they had not done fetal histopathology, but felt it would be worth looking at in future studies.

Dr. Boekelheide addressed the issue of common pathways. He agreed that if you look at an exposure at a specific instant in time, there will be shared effects regardless of the mechanism. However, each mechanistic pathway may cause differing physiological alterations in the tissue after a longer term exposure. Dr. Gray agreed that it was a valid point, but stated that the studies were only over five days in development of the fetus, and he was not sure there would be effects much earlier or later, and he did not know that the fetus had much capacity for compensation. Dr. Foster followed up by saying that his group looked at the fetal epididymis and androgen receptor expression following treatment and found that there is a similar pattern of down regulation with phthalates as with androgen-receptor antagonists. Phthalates were not affecting the epididymis *per se*, but rather the epididymis was affected as a consequence of the lowered androgen levels, but the effect is quite similar to what one gets when blocking the androgen receptor.

Dr. Borgert then revisited the question of dose addition versus response addition in cumulative risk assessment of phthalates. He felt that the ability to predict the dose-response using dose addition would be very tenuous at lower doses. Dose addition assumes that chemicals are acting as dilutions of one other, which requires a common endpoint of measurement. In the case of low doses of phthalates and low doses of androgen receptor antagonists, the antagonists will block some receptors, but you will still have plenty left for what testosterone is there, so the two chemicals might be acting independently at very low doses, say significantly below the range of observed effects. It is only when you get close to the functional reserve capacities of those pathways that you start to see additivity. Dr. Gray emphasized that researchers should not abandon the concept of common mechanism of toxicity and it should be included in risk assessment, but said that they are trying to broaden it. He added that the dilution-addition principle would hold for the types of chemicals talked about so far because the mode of action at the level of the gene is identical. At the level of the gene, it is one thing that the tissue sees. At small doses and many chemicals, it can be difficult to detect things; you have very small concentrations of a lot of chemicals and we are trying to predict 1 percent change. But this point does need to be discussed more. There was further discussion regarding the potential overlap of response addition and dose addition at very low doses. Dr. Gray explained that this overlap, where one cannot distinguish between dose addition and response addition, often occurs with chemicals that have linear dose-response curves, but for chemicals with a steep threshold, the response and dose addition predictions will be different.

Most panelists agreed that in most cases, dose addition is more protective and predictive than response addition.

Dr. DeVito suggested that if we were doing a cumulative assessment for six chemicals, even up to 12 chemicals, each one would have to be at almost a tenth of the NOAEL, so you are not really extrapolating that far down the curve. But if what we are using is an RfD, it does not matter what all those really low exposures add up to, as long as they do not come close to the RfD where something happens. Dr. Borgert pointed out that, actually, the RfD is already a hundred to a thousand times lower than the NOAEL because of various safety factors used to construct it. If we were using the NOAEL, dose addition would be appropriate, but several orders of magnitude lower makes it difficult to determine whether dose addition or response addition is appropriate. Dr. DeVito asked whether that meant that if he were doing a cumulative risk assessment for the chemicals, it would make a difference when the addition took place, before or after the uncertainty factors, especially for human extrapolation.

Dr. Borgert stated that that was a good point, and that here we have the opportunity to use human data (which would make the uncertainty factor unnecessary). He noted that when human data are available, they should be used instead of applying uncertainty factors to extrapolate from animal data. He brought up the case of diethylstilbestrol (DES) and the data available on the thousands of women exposed and the fetal reproductive tract malformations that occurred in their male offspring. Dr. Borgert indicated that one can get a human potency relevance threshold from these reproductive data, which is more informative then applying several uncertainty factors to animal data. Other panelists, however, countered that the DES effects were not the same as those seen in the phthalate syndrome and that they occurred by a completely different mechanism; thus, DES comparisons should not be made.

Ms. Teuschler concluded this part of the discussion by summarizing that dose addition has two applications. It can be applied to toxicological studies when trying to be accurate, and it can also be applied in doing a mixtures risk assessment in order to be protective (this is where the uncertainty factors come into play). There may be other ways of doing a mixture assessment in which one might want to look at relative potency factors. If one does use human data, one will still have to consider uncertainties and variability.

Dr. Foster summarized by saying that it would be a good idea that if doing a dose-response assessment for phthalates, one should consider the phthalate syndrome as a whole, but that it would be difficult to do from literature-based studies if you did not have access to all of the data, because not all studies measured or recorded all of the effects of interest in the syndrome. Therefore, it is important to be able to access and evaluate the quality of the raw data. Dr. Gray commented that there were up to 10 phthalates in his studies producing positive results; does this mean the list of six should be expanded? You can put 100 chemicals into your analysis, but, generally, all that means is that you have a larger spreadsheet; for 95 of them, this particular population does not have any exposure. It is not that you did not consider it; rather, it just is not contributing anything. For example, for DPP among the six, there does not appear to be high exposure now, but because it is in the analysis, if it were to be used as a replacement for something else, you would have a way to show its predicted impacts. So if you have nine things in your equation, you may find that there may be problems only for certain parts of the population. If the analysis shows no problem, then great, we can tell people we did this analysis; if there are problems, then EPA can do something.

Dr. Foster asked Dr. Gray what the most sensitive outcome is. In terms of endpoints, Dr. Gray hypothesized that the cascade of effects should be the same for all of the phthalates and that reduction in fetal testosterone will always be a more sensitive endpoint than the phthalate syndrome; epididymal agenesis will always occur at a lower rate than hypospadias; AGD will not be the most sensitive. But

various researchers have done things differently, so it might not always fall out that way in the literature. Dr. Foster asked if they know the POD for the fetal testosterone. Dr. Gray indicated that they have some data on dose-response with fetal androgens and malformations. He noted that finding a POD is doable, and they are trying to do it. The hypothesis is that with a 50 percent reduction in testosterone with any phthalate or mixture of phthalates, one will see the same profile of effects.

Dr. Rider suggested that it would be good to rely on the individual phthalate data rather than the mixtures data, so that new phthalates can be added or removed (if they are banned) in the future. Dr. DeVito asked whether one could use a relative potency approach to predict effects on fetal testosterone, even for chemicals for which one did not have data on fetal testosterone, but did have other syndrome dose-response data. Dr. Gray was not sure if that was necessary because the fetal testosterone data is mostly complete, and for chemicals for which it is not, it will likely be collected. The assumption, however, that one could use other syndrome data to predict fetal testosterone effects is reasonable.

Dr. Gennings inquired if there were data available for substitute phthalates or substitute chemicals replacing traditional phthalates. Dr. Gray said that there are some substitutes becoming more popular, like isobutyl phthalate, which has been studied. Di-propylheptyl phthalate has no published literature, even though it is becoming more popular as well. DINCH[5] (a phthalate substitute) is assumed to be negative for reproductive effects. Dr. Gray continued on to say that as substitutes arise, you would like to have the toxicity data, and if substitutions do not affect the same pathway as the current phthalates being assessed, it would not be added to a phthalate cumulative risk assessment. If it has unique toxicities of its own, it would have to have a separate assessment.

Dr. Foster concluded this portion of the animal data section by recapping the four questions posed by EPA. The first was whether EPA could use the phthalate syndrome as a whole for characterizing cumulative hazard. Panelists agreed that it would be beneficial, but that it may not be possible even if the raw data was obtained because in some studies, the data were not collected. Secondly, the EPA asked if the assessment would be based on the most sensitive outcome and what that outcome is. The panel responded by saying that fetal testosterone appears to be the most sensitive endpoint based on the data, but that gene expression changes are synonymous and gene expression changes are more reproducible with the newer platforms. Finally, EPA asked if one would base the assessment on mixtures data or individual data. The panelists felt that one could base it on both individual and mixtures evaluations. Dr. Foster concluded by noting that EPA should not lose sight of the non-androgen-mediated endpoints (see next section); some of these may need to be considered for inclusion as sensitive endpoints.

3.5. Animal Data for Phthalates: Other Toxicities Associated with Phthalates, Dr. Jason Lambert (EPA/ORD)

Dr. Lambert continued the topic of using animal data (including mixtures data) for phthalates for conducting a cumulative hazard assessment by presenting information on other health effects of phthalates (i.e., excluding developmental and reproductive outcomes). Dr. Lambert noted that there are common outcomes between phthalates for the liver, kidney, hematological system, thyroid, immune system, neurological system. He noted that body weight effects have also been observed following exposure to phthalates.

In terms of effects on the liver, although there is no liver data in humans, there is significant evidence of effects on the liver from studies in other mammals. Rats and mice are considered relatively sensitive,

[5] DINCH is 1,2-cyclohexanedicarboxylic acid, diisononyl ester.

with male rodents more susceptible than females. Dr. Lambert noted that for some phthalates, the liver effects are occurring within the range or at the same dose level as the reproductive/developmental effects. The most common liver effects include increased organ weight, enzyme activity alterations, and peroxisome proliferation. Several functional pathways and organelles are affected (e.g., lipid and carbohydrate metabolism, microsomal mixed-function oxidase system, and mitochondria).

Nephrotoxicity has also been observed following exposure to a variety of phthalates including DEHP, DBP, BBP, DINP, DIBP, and DPP. Renal effects, such as organ weight changes, have been observed in rats and mice, but not other laboratory species, and it is unclear how relevant the data are for assessment purposes. For example, some phthalates caused increased kidney weight (e.g., DBP, BBP, and DINP), while others (e.g., DEHP, DPP, and DIBP) caused decreased kidney weight. Less common renal effects associated with phthalate exposure include inflammation and pigmentation of tubular epithelium. As in the liver, some of the renal effects are occurring within the range or at the same dose level as the reproductive/developmental effects.

Dr. Lambert continued by discussing the potential for carcinogenicity from exposure to phthalates. Although various phthalates are associated with cancer, the cancer type is not consistent among phthalates. DINP appears to have the most evidence of carcinogenicity, with an increased incidence of mononuclear cell leukemia in male and female rats and kidney neoplasia in male rats exposed to this phthalate. No carcinogenicity studies for DBP, DIBP, or DPP were identified.

Dr. Lambert concluded his presentation by reemphasizing that the six phthalates selected for the cumulative hazard assessment have some common non-reproductive/developmental outcomes. While some effects are conserved across outcome (e.g., liver weight), others are differentially expressed across the phthalates (e.g., kidney weight changes). Finally, there are a few additional conditions, including insulin resistance and increased adiposity, which have not yet been well characterized.

Discussion Summary

Dr. McIntyre asked Dr. Lambert what the dosing paradigm was for most of these studies. Dr. Lambert clarified that the studies were subchronic or longer, so animals were dosed for 90 days or more. In addition, most of these studies dosed young adult animals.

Dr. Boekelheide commented that at least some of the phthalates are PPAR (peroxisome proliferator-activated receptor) agonists, so hepatomegaly would be expected at some dose based on the pathway. However, Dr. Boekelheide pointed out that this pathway is less relevant for humans. He also noted that there is an upcoming NIEHS workshop on obesogens in January, and that he expected phthalates to be part of the discussion.

Dr. Foster then asked if the reproductive and developmental changes had any relationship to PPAR alpha. The panelists did not think they were related. Dr. Gray indicated that they found consistent and measurable amounts of PPAR alpha in fetal testes, but no gamma. However, they did not see any indication that that pathway was activated. Dr. Boekelheide added that this would be consistent with other pubertal studies conducted on PPAR alpha knockout mice.

Ms. Teuschler of EPA/ORD asked if there are animal data on routes of exposure other than oral, since human exposure is presumably multi-route. Dr. Lambert indicated that there are few data on routes other than oral exposure, although DEHP has some inhalation data. Dr. Lioy and Dr. Gray suggested that the dermal route also needs to be considered based on some of the cosmetic and personal care product exposure in humans. If phthalates are in house dust, one would ingest it before inhaling it.

3.6. Summary of Day 1

Based on the presentations and discussions throughout the day, Drs. Lambert and Hotchkiss identified and presented the following summary points. Since the goal of this workshop was to collect individual input rather than achieve consensus, the group did not discuss these points in detail; rather, they reviewed the list in its entirety to ensure that additional points should not be included.

Summary Points for Day 1:

- Dose addition has been shown to perform well with mixture data and is considered more predictive and protective than response addition methods.
  - One panel member was concerned that dose addition may not be applicable at certain very low dose levels.
- Epidemiologic literature indicates that humans are exposed to phthalates, and there is limited evidence that phthalate metabolites are associated with reproductive effects in the developing fetus (males and potentially females) and adult males.
- Epidemiologic studies on phthalates can be improved by focusing on the relevant effects and improving methods for characterizing relative source contribution.
- A systematic approach for reverse PK/TK methods is needed for multiple routes of exposure, in particular oral and dermal.
- Fetal testosterone appears to be the most sensitive endpoint based on the data, although gene expression changes are similar.
  - Non-androgen dependent outcomes and non-reproductive/developmental outcomes should also be considered.
- Panelists agreed that EPA could base the assessment on both component-based and whole mixture approaches.

4. Day 2 – Methods for Conducting the Cumulative Assessment of Six Selected Phthalates

Day 2 of the workshop focused on exploring the different methodologies available for conducting a cumulative assessment of the six selected phthalates. Specifically, the presentations on Day 2 looked at how the human and animal hazard data discussed during Day 1 could be applied to mixtures modeling approaches. Discussions following each presentation allowed the invited panelists and discussants to: (1) consider the strengths and limitations of each method given the available hazard data for the phthalates of interest; and (2) identify assumptions or uncertainties that are associated with each method.

Ms. Linda Teuschler presented on cumulative assessment approaches for multi-route exposures to phthalate mixtures, including the derivation of cumulative relative potency factors (RPFs), while Dr. Chris Gennings provided presentations on the hazard index (HI) method and the point of departure index (PODI) method, which stemmed from the 2008 NAS recommendations. Dr. Gennings also presented a novel biomonitoring based approach that is currently under development.

Dr. Cynthia Rider provided the final presentation of Day 2, a case study using phthalate mixture data, which included incorporating other anti-androgens, in addition to phthalates, to conduct a cumulative assessment.

4.1. Candidate Risk Assessment Approaches for Multi-Route Exposures to Phthalate Mixtures, Ms. Linda Teuschler (EPA/ORD)

Ms. Teuschler provided an introduction to chemical mixture risk assessment theory and component-based methods. She began by reviewing the components of the risk assessment paradigm—hazard identification, dose-response assessment, exposure assessment, and risk characterization—and noted that, in addition to the risk assessment issues inherent to evaluating single chemicals, the application of this paradigm to mixtures requires consideration of an expanded set of issues (see Figure 6). For example, additional issues for hazard identification may include identifying the effects from the total mixture dose and the potential effects that may result from interactions. The dose-response step may also involve considering the potential for effects below individual chemical thresholds and incorporating judgment regarding similar toxicity within or between chemical mixtures. In exposure assessment, one may need to account for multiple exposure routes and pathways, as well as internal dose of mixture components at the target tissue. She noted that exposure assessment and dose-response are interdependent, as variations in a mixture's components, dose levels and proportions will impact the nature of its dose-response. Ms. Teuschler continued that risk characterization includes evaluating the data and data gaps, and considering uncertainty of changes in exposure caused by chemical-to-chemical interactions during fate and transport in the environment. Finally, the goal of the assessment must be clear when selecting a risk assessment method. For example, a risk characterization may involve calculating a HI, which could be used for several purposes, such as, to inform regulations, to set safe levels of exposure, or to prioritize research.

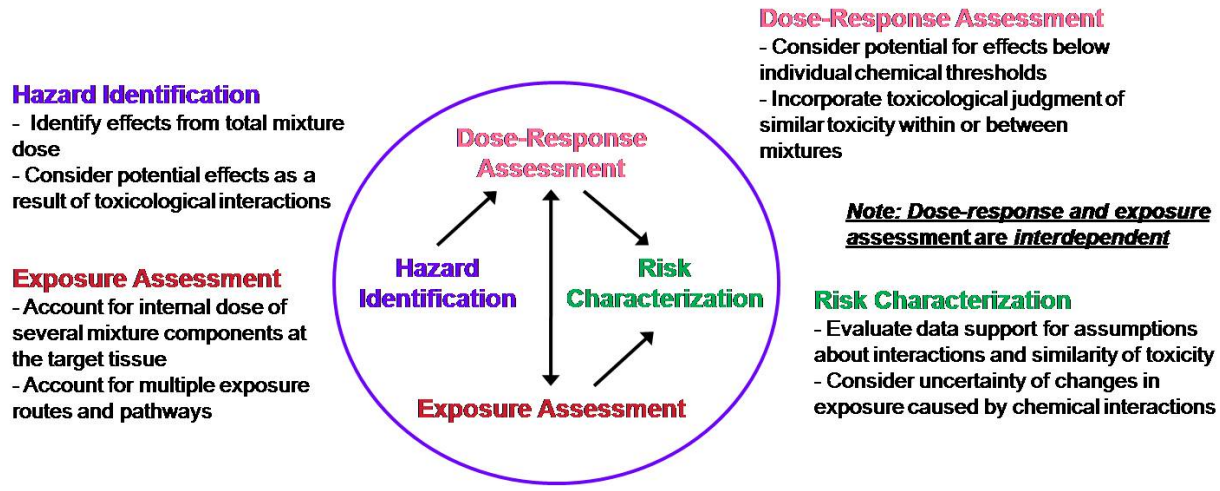

Figure 6. Risk Assessment Paradigm for Chemical Mixtures: In Addition to Issues for Single Chemicals

Ms. Teuschler went on to discuss various models of additive joint toxic action of mixture components. She focused mainly on dose addition, a type of simple similar action, which involves the addition of component doses scaled for relative toxicity (e.g., HI, cumulative HI [CHI], RPFs, mixtures reference value [RfV], mixtures BMD, mixtures margin of exposures). She described another method, response addition (considered simple dissimilar action), which adds component risks and assumes toxicological and statistical independence (e.g., cancer risk sums). A third model, integration of dose addition and response addition (e.g., cumulative RPFs) is considered when the mixture contains chemicals with mixed toxic modes of action. This method involves the addition of risks calculated for common mode of action chemical subgroups, assuming toxicological similarity within subgroups, but independence of toxic action between subgroups.

Figure 7 presents a continuum of knowledge regarding chemicals, with the highest level of knowledge being at the mechanism of action level. For example, with dioxins it is well known that they bind to the Ah receptor. Typically, however, a chemical's mechanism of action is unknown, so the next level would be to look for information about key events along the toxic mode of action. At this level of knowledge general methods may be used that may be limited by restricting the assessment to a specific route, endpoint, or exposure duration. An example of applying a similar mode of action method is the calculation of RPFs for a specific effect. If one does not have information on mechanism or mode of action, it is possible to use information at the target organ level to complete a screening level assessment (e.g., developing a risk indicator such as an HI).

Factor	Information on Toxicological Action		
Terminology	Mechanism-of-Action	Mode-of-Action	Toxicological Similarity
Level of Knowledge	High	Medium	Low
Likelihood of Knowledge	Low	Medium	High
Knowledge of Toxicological Action	At the Cellular/ Subcellular Level	At the Tissue level	At the Target Organ Level
Example Toxic Effect & Real World Mixture	Dioxins (Binding to Ah Receptor) (e.g., van den Berg et al., 2006)	Organophosphorus Pesticides (Cholinesterase Inhibition) (e.g., U.S. EPA, 2002)	Contaminated Site Assessment (Various Liver Effects) (e.g., U.S. EPA, 1989)
Choice of Risk Assessment Method	Specific methods, e.g., TEFs*	General Methods, Limited by Route, Endpoint, Exposure Time, e.g., RPFs	Screening Level Assessments, e.g., HI, Cumulative HI

*Toxicity Equivalence Factors

Figure 7. Similarity of toxic action: As level of knowledge increases, the likelihood of having that knowledge decreases.

Ms. Teuschler next discussed the CHI method, which accounts for multiple chemicals and multiple exposure pathways using dose addition. To develop a CHI, the first step is to develop a unitless hazard quotient (HQ) for each single chemical, for each exposure route (Figure 8a). Following this step, there are two approaches to calculating the CHI that provide slightly different information to the risk assessor. In the first approach, these HQs are first aggregated (summed) across the 'm' exposure pathways for each chemical (Figure 8b), which provides an estimate of the chemical with the greatest risk potential regardless of exposure pathway; then, one can sum the aggregated HQs across the 'n' chemicals to calculate the CHI for the mixture (Figure 8c). The second approach is to first sum across the 'm' exposure pathway-specific HQs, which provides an estimate of the exposure pathway with the greatest risk potential regardless of the chemicals and then to sum across the 'n' chemicals to calculate the CHI for the mixture (Figure 8d). Both approaches result in providing an indication of potential health risk from multiple-pathway exposures to the chemical mixture.

a) $$HQ_{jk} = E_{jk} / RfV_{jk}$$ *Where *RfV* = risk based value (e.g., RfD) same units as Exposure, E

b) $$HQ_j = \sum_{k=1}^{m} HQ_{jk}$$ *For j^{th} chemical and k^{th} pathway

c) $$CHI = \sum_{j=1}^{n} HQ_j$$

d) $$CHI = \sum_{k=1}^{m} \left(\sum_{j=1}^{n} HQ_{jk} \right)$$

Figure 8. Hazard quotient calculations: a) Pathway-specific hazard quotient for chemical j and pathway k, say, a phthalate in drinking water, b) Aggregating HQs for the single chemical j over all (k = 1 to *m*) pathways, c) CHI as sum of HQs for a mixture of all (j = 1 to *n*) chemicals, d) CHI as sum of all (k = 1 to m) exposure pathway-specific HIs.

Ms. Teuschler then discussed the mixtures BMD approach and how to calculate a mixtures BMD for a single exposure route/pathway using a rearrangement of Berenbaum's equation (Berenbaum, 1989). This approach is restricted by considering the fraction of each chemical in the mixture dose, which is limited by the proportions used in the experiment (see Figure 9). The fraction can also be calculated for environmentally relevant proportions.

$$ED_{xm} = \frac{1}{\sum_{i=1}^{n} f_i / ED_{xi}}$$

Figure 9. Mixtures benchmark dose for a single exposure route/pathway, where: ED_{xm} = effective dose at x percent response for the mixture (e.g., an ED_{10} for the mixture). In the denominator on the right side of the equation is the expression f/ED, where f_i = fraction of mixture dose represented by chemical *i* in the experiment, which is divided by its individual effective dose (i.e., ED_{10}, or ED_{xi}). These individual component f/EDs are then summed over all components, where n = number of components in the mixture

She noted that in order to develop a mixtures reference value, one can use a similar equation (see Figure 10). Like the equation in Figure 9, this equation is also suitable for fixed or similar proportions of chemicals in a mixture and can be calculated for a range of environmentally relevant proportions. But since now we are using RfVs instead of EDs, one must consider factors that can impact uncertainty including: how stable the exposure is, how good the dose-response information is, and what the uncertainty associated with the calculated reference value is, since the uncertainty factors used to calculate the RfVs for each chemical component may be different.

$$RfV_m = \frac{1}{\sum_{i=1}^{n} f_i / RfV_i}$$

Figure 10. Mixtures reference value, where RfV_i = reference value of chemical *i* , f_i = fraction of mixture dose represented by chemical *i*, and n = number of components in the mixture. These calculations can be made either for specific chemical components of a mixture or specific routes/pathways.

Ms. Teuschler then discussed the cumulative RPF method, which is an approach that integrates the RPF method with response addition. In this case, the mixture is subdivided into two or more common mode of action subgroups and, for each subgroup, a risk estimate is developed assuming dose addition via an RPF approach. The subgroup risks are then added assuming toxicological and statistical independence via response addition.

In the RPF method, an "index chemical" among the mixture components is chosen, and the doses of each of the other components is adjusted to an index chemical equivalent dose (ICED) using a RPF (RPF for the component chemical), so, under dose addition, they can be added to estimate an ICED for the mixture. The RPF, or scaling factor, is generally calculated either as the ratio of the ED_{10} (effective dose at which a 10 percent response is observed) of the component to the ED_{10} of the index chemical, or the ratio of the potency of the index chemical to that of the component chemical (i.e., ratio of their slope factors). Uncertainties with this method include both index chemical uncertainties (How good are the toxicological data for the index chemical?) and mixture uncertainties (How similar are the other chemicals to the index chemical? How much of the mixture is composed of the index chemical?).

The statistical theory for response addition uses the statistical law of independent events, which states the mixtures risk for a n-component mixture is 1 minus the probability of not responding to any of the mixture component chemicals (see Figure 11a).

a) $$Rm = 1 - \prod_{i=1}^{n} (1 - r_i)$$

b) $$R_m = 1 - (1 - r_1) * (1 - r_2)$$

c) $$R_m = r_1 + r_2 - r_1 * r_2$$

Figure 11. a) The statistical law of independent events for exposure to a mixture of n chemicals; where Rm = mixtures risk and r_i= % animals responding from exposure to chemical *i*, b) The statistical law of independent events for exposure to a mixture of 2 chemicals, and c) Equation b is algebraically shown to be the sum of the component risks minus their product and can be interpreted as 1 – (probability of not responding to either chemical).

Figures 11b and 11c illustrate the method for a 2 chemical mixture, where 11c is algebraically equivalent to 11b. In this case, when the probability of response is small, as is the case for many environmental exposures, the multiplicative term on the right hand side of the equation in Figure 11c becomes very small, and the overall risk is well approximated by the addition of the two probabilities of response.

It may be noted that dose addition and response addition provide risk estimates that may be interpreted differently because of the assumptions of common and independent toxic modes of action

across mixture components, respectively. For example, if doses related to liver cancer by a specific mechanism were being added under dose addition, the resulting risk estimate would relevant to liver cancer via that mechanism. In contrast, if risks for several independent types of cancer were being added via response addition, the resulting risk estimate would be relevant to the expression of "some non-specific form of cancer".

As previously noted, cumulative RPFs are calculated by integrating dose addition and response addition. This approach could be used in the case where the chemicals of interest have the same health outcome, but there is one subgroup of chemicals with one defined mode of action and another subgroup of chemicals causing the same effect via another mode of action. To calculate cumulative RPFs, for each mode of action subgroup, one sums the individual component ICEDs to generate each subgroup's ICED, and then calculates that subgroup's risk. Then the subgroups risks are added together. The resulting "integrated" risk would be interpreted as the risk of some form of the health outcome of interest that is common to both subgroups.

Ms. Teuschler concluded that there are a number of uncertainties surrounding component-based approaches. In general, these approaches should be limited to simple, defined mixtures, and even then the assessor will be forced to use considerable judgment. She indicated that it is important to identify and confirm assumptions for all of the approaches presented, where possible, and be transparent in describing all data gaps, data quality, and data quality differences among chemicals.

Discussion Summary

The panelists and discussants addressed the various methods and approaches that Ms. Teuschler presented, including their strengths and weakness.

Dr. Gray began the discussion by explaining his group's attempts at separating chemical groups by differing modes of action and performing integrated addition. He noted that they separated seven phthalates and ten anti-androgens into separate modes of action and used dose addition for each group independently and then used response addition to sum for the integrated risks. They found that this was not useful for prediction, as putting them into different mode of action groups yielded too low of a prediction compared to the observed effect.

Dr. Borgert commented on the fact that both the dose addition model and the response addition model are mathematical constructs that exist independently of biology. One can input a value into either model and get an answer, but it is up to the analyst to justify why the particular model used makes logical sense biologically. The critical issue is deciding what parameters will allow one to choose a model that is most predictive. Ms. Teuschler responded to Dr. Borgert by saying that to use response addition, the processes need to be truly independent. In addition, although dose addition is fundamentally a mathematical construct, scientists should use dose addition when it makes biological sense to do so. She also clarified that these are very simple models for very complex biological processes.

Dr. Foster further questioned if you have empirical evidence that chemicals follow dose addition, even though the mechanisms of action are different, would you use dose addition models? In this case, integrative addition models might not be as predictive as one might have thought. Dr. Borgert stated that if you truly have an almost identical mode of action, you can have confidence that your prediction at lower doses will stand up. As they get more and more mechanistically dissimilar, one becomes less confident in the ability to extrapolate to lower doses and other species, even if dose addition works at the observable effect level.

Dr. Foster then asked the panelists to revisit the relative strengths and weaknesses of the different approaches. He asked if in using a Toxicity Equivalence Factor (TEF) approach, the same RPFs are used

whether it is a cancer assessment or a developmental assessment. Dr. DeVito commented that for the TEF approach, scientists have always been challenged by predicting endpoints for chemicals for which there is no data, so they have used surrogate relative potency. Dr. Foster commented that any of the methodologies described by Ms. Teuschler could be relevant, and that any and all of the methods might be helpful for what's being proposed here; therefore, one cannot throw any of them out.

Dr. Borgert raised concern again over whether one can know if by using all of these new (cumulative) methods if they've made the risk assessment more accurate, or whether they've introduced more uncertainty with these methods, versus traditional organ-based hazard index approaches. When empirical data exists we can confirm accuracy, but we do not always have data. Dr. Gennings replied that in general, there are very good mixtures data available for phthalates. Dose addition worked well with the data available and that is why the NRC panel recommended this approach. The hazard/point of departure index approach attempts to predict effects in animals and then compare. When there are good animal data, then dose addition works well in extrapolating to humans.

Dr. Boekelheide commented that the database is rich for phthalates and related chemicals so one could create, within a species, a scale of what the endpoint biomarkers were, relative to each other. One could make a consolidation of endpoints. However, there is not much known about scaling between species, for example, rats to humans. Dr. Foster agreed that if one cannot adopt some sort of mixtures approach for phthalates, one is not going to be able to do so with other chemicals with less complete data sets.

Dr. Gray spent some time discussing uncertainty of phthalate data in humans. Although there is a lot of data on the human-to-animal scaling for pharmaceutical estrogens and other pharmaceuticals, that type of information does not exist for phthalates. Dr. Gray commented that researchers do not know with 100 percent certainty that the phthalate syndrome would ever happen to humans at all. However, there is an assumption that the pathway is relevant until proven otherwise. One has to live with uncertainty and move on with a risk assessment.

Dr. DeVito then asked about whether the scientific community understands the relationships of scale between rats and humans. With the phthalate syndrome, it is unclear by what mechanism fetal testosterone is decreased. It appears that there is no quantitative data in humans that allows us to know how much testosterone has to be decreased in humans to see an effect similar to what is seen in rats. Dr. Fox advocated for an approach in the coming assessment that was presented by Dr. Cooper on Day 1, which uses toxicokinetic models for particular phthalates to back calculate the exposure to the parent compound. This type of analysis would not necessarily result in better uncertainty factors, but would at least give an idea of the sensitivity of humans relative to animals based on exposure to the parent compound. She noted that this would at least put some data behind the scaling.

Dr. Foster concluded this section of the workshop by saying that there is not going to be complete harmony on what approaches to take, but none of the approaches can be ruled out.

4.2. Hazard Index Method, Dr. Chris Gennings (VCU)

Dr. Gennings described the steps for conducting a cumulative risk assessment using the HI for combined exposures to anti-androgenic chemicals, including phthalates. Her presentation was based on the study entitled, "Combined exposures to anti-androgenic chemicals: steps towards cumulative risk assessment" (Kortenkamp and Faust, 2010), as well as the 2008 NAS phthalate report (NRC, 2008). Dr. Gennings began by reviewing the ubiquity of human exposure to androgen disrupting substances, including phthalates. She emphasized that reducing human exposure to these chemicals would not be a simple process, but would require many changes across the board. She discussed the importance of androgens

in fetal development, noting that any suppression of androgens through inhibition of steroid synthesis or by antagonism of the androgen receptor could cause irreversible de-masculizing effects later in life. She also mentioned two publications that support a possible association between anti-androgen exposure and decline in semen quality (Sharpe, 2009) and other de-masculinizing effects (Swan et al., 2005) in males. In addition, Dr. Gennings cited guidance from the 2008 NAS report (NRC, 2008), which advised that risks associated with phthalates should be evaluated by taking into account combined exposure, and should include other chemicals that cause similar health effects.

In order to demonstrate the value of the HI as a method of cumulative risk assessment and recognizing that the 'background' exposure to anti-androgens occurs from many sources, Dr. Gennings evaluated the combined exposure to phthalates by calculating two hazard indices for the same group of 15 chemicals, using two different sets of assumptions. Eleven of these chemicals (five phthalates- DEHP, DnBP [di-*n*-butyl phthalate], DIBP, BBP, DINP and six other chemicals- vinclozolin, prochloraz, procymidone, linuron, fenitrothion, p,p'-DDE) were selected based on substantial *in vivo* and *in vitro* evidence of their ability to disrupt male sexual differentiation in reproductive toxicity models in rats (i.e. exhibited effects characteristic of the androgen insufficiency syndrome). Four of these chemicals (BDE 99, bisphenol A, butyl paraben, propyl paraben) were selected based on limited *in vitro* evidence and weaker *in vivo* evidence of androgen antagonism. She continued by explaining usages and common endpoints for these chemicals.

Dr. Gennings next explained the concept of dose addition and how it was applied to these hazard index calculations. She indicated that dose addition is not limited to mixtures with the same mechanism of action and can be applied to substances with different dose response slopes. In this case, she noted that mixtures of phthalates with other anti-androgens may exert their effects through different molecular mechanisms; however, they may cause the same phenomenological effect (e.g., human testicular dysgenesis syndrome; NRC, 2008). Therefore, she explained, the dose addition principle can be used to approximate the combination effects of phthalates with other anti-androgens.

Dr. Gennings then explained the HI method as the sum of hazard quotients, defined as the ratio of exposure (e.g., daily intake) to an acceptable level for a specific chemical for the same period of time (see Figure 12). She stated that the difficulty with the hazard index was in finding the right values to input into the equation, namely defining RfDs and daily intake (DI) values.

$$\text{Hazard Index (HI)} = \sum_{j=1}^{c} \frac{\text{DI}_j \ (\mu\text{g/kg/day})}{\text{RfD}_j \ (\text{AA}; \ \mu\text{g/kg/day})}$$

Figure 12. Hazard index equation; where c is the number of chemicals, DI is the estimated daily intake, and RfD is the reference dose for the anti-androgenic (AA) chemical, defined by *in vivo* evidence of anti-androgenic effects.

Next, Dr. Gennings worked through the hazard index equation for two cases using different RfDs. Case 1, adapted from Kortenkamp and Faust (2010), used RfDs determined using anti-androgenicity data to estimate the PODs—typically based on lower limits of benchmark doses for the phthalates and NOAELs for the other anti-androgens. She compared these results with a second case, in which she assumed that DIBP, DnBP, DEHP and BBP were approximately equipotent. For DINP, she used a higher RfD based on relative potency. She used the same RfDs for anti-androgens as in Case 1.

Intake level estimates were obtained from peer reviewed literature or from publicly available reports from the European scientific community, and Dr. Gennings cautioned that exposure may differ between Europe and the United States. Dr. Gennings also noted that uncertainty differences varied among

phthalates in Case 1. For DnBP, DIBP, and BBP the uncertainty factor was 200, due to small study sample sizes. The uncertainty factor for DINP was 500 because the RfD was calculated from the LOAEL rather than the NOAEL. DEHP had an uncertainty factor of 100. All of the phthalates in Case 2 had an uncertainty factor of 100.

Dr. Gennings then presented the values she obtained for Cases 1 and 2 using median intake levels. Despite the different assumptions made for each case, they led to very similar hazard quotients—0.38 for Case 1 and 0.39 for Case 2. She identified the substances that were main contributors (i.e., contributed to at least 5 percent of the HI) for both Cases 1 and 2 as DBP, DEHP, vinclozolin, prochloraz, procymidone, and butyl paraben. DiPB was classified as a large contributor in Case 2 (7.72 percent), but not in Case 1 (2 percent).

Next, Dr. Gennings calculated the HIs for Cases 1 and 2 using the high intake values, rather than the median values. She acknowledged that these high intake numbers did not portray real-world exposure levels for these compounds, but introduced it as a worst-case scenario. For this calculation, the hazard indices for Cases 1 and 2 were again quite similar, at 2.01 for Case 1 and 2.12 for Case 2. Using high intake values, the main contributors to the HQ for Cases 1 and 2 were vinclozolin, prochloraz, bisphenol A, and butyl paraben. DBP was a large contributor (5.66 percent) for Case 2, but not for Case 1 (2.98 percent), while DEHP was a large contributor for Case 1 (5.94 percent), but not Case 2 (3.40 percent).

Finally, Dr. Gennings calculated the contribution phthalates made to the HI in Case 1 and 2 at median and high intake values. When median intake levels were used to represent exposure, phthalates contributed 32 percent to the HI in Case 1 and 34 percent in Case 2. When high intake values were used to represent exposure, phthalates contributed 10 percent to the HI in Case 1 and 14 percent in Case 2.

A summary of the results presented by Dr. Gennings is presented in Figure 13. Dr. Gennings concluded by reviewing the following assumptions she made during this analysis:

- Joint action of anti-androgens can be approximated by dose addition
- Human populations are exposed to each of the chemicals included in the case studies and the estimated intakes are relevant to pregnant women
- Selection of anti-androgens was driven by chemicals with available *in vivo* data – additional chemicals with known human exposure are likely to contribute to cumulative anti-androgen risk
- Important to evaluate risk from phthalates while accounting for other anti-androgens

Hazard Index (% from phthalates)	**Median intake estimates**	**High intake estimates**
RfD Case 1: (Kortenkamp and Faust, 2010)	0.38 (32%)	2.01 (10%)
RfD Case 2: Assuming equi-potency (Gray, 2008)	0.39 (34%)	2.12 (14%)

Figure 13. Summary results for the HI case studies.

Discussion Summary

The panelists and discussants asked some clarifying questions, as well as provided EPA with some suggestions on next steps for incorporating this approach.

Dr. Borgert began the discussion by asking Dr. Gennings to clarify her statement that she had "broadened the definition" of dose addition. Specifically, he was unclear as to whether she was using a new mathematical model to calculate dose addition, or if she was applying the traditional mathematical model to a broader set of chemicals. She clarified that she was not changing the mathematical model or the definition, but rather than limiting the dose addition method to chemicals with the same mode of action and parallel dose-response curves, she was applying the model to a set of chemicals with different modes of action and varying dose response curves. The 2008 NAS committee on phthalates was approaching it by looking at phenomenological effects that could be associated with a number of different biological pathways, and trying to be more environmentally relevant. Dose addition is a mathematical concept, but the limitations of parallel dose-response curves and the same mode of action are not restrictions in the definition, so we're applying it on a broader scale than previously described.

Ms. Tueschler voiced concerns on the validity of the exposure estimates and the reference dose calculations in the two cases outlined by Dr. Gennings. She opposed using reference doses with uncertainty factors higher than 100, did not know of any cases where EPA used an uncertainty factor of 200, and thought that these would have to be recalculated if EPA were to use this approach. She proposed that the panelists devise some criteria with which to advise EPA on what chemicals to look at together for a cumulative risk assessment. She continued that in the case of expanding the six phthalates, EPA may not do that immediately, but some suggestions about what additional chemicals to include and when to include them would be welcomed. Dr. Gray recommended looking at chemicals with reasonable *in vivo* evidence, as *in vitro* studies may provide misleading results. Dr. Foster noted that when *in vitro* data is used and the chemicals become drivers in the risk assessment, you may have a problem.

Dr. DeVito asked for clarification on the definition of an anti-androgen, and recommended that the panelists devise a solid definition and specific endpoints to monitor. He recommended that EPA devise a number of RfDs for phthalates for different endpoints, rather than a single RfD for all outcomes, so when they develop a cumulative assessment, they could group the RfDs for specific effects. Although some of these reference values might not be based on the most sensitive effects across the selected phthalates, they could be used in the context of a "bucket" of effects that all fall in the category of phthalate syndrome, and would inform the assessor whether the chemical should be included in the assessment or not. For example, fetal testosterone levels may be a good endpoint to base a POD for phthalates, while AGD measurement may be a better indicator for other anti-androgens. This way, cumulative risk could be assessed for a number of specific endpoints. There was general consensus that this was a good idea. Dr. Fox pointed out that this application of cumulative risk could be expanded to other classes of chemicals beyond phthalates, and this was actually one of the recommendations the NAS made in their report (NRC, 2008). Dr. Gray pointed out that when you start including other anti-androgens (e.g., androgen receptor antagonists) in the assessment, you can no longer use fetal testosterone depression (or the gene surrogate) as the most sensitive effect, because although they may cause the same ultimate phenomenological effects, they may have no effect on testosterone levels. You would have to identify PODs from among the suite of effects associated with anti-androgenic activity.

Dr. DeVito summarized by saying that if one was doing a cumulative assessment, one might select an effect from the suite of malformations associated with decreased fetal testosterone in identification of a POD for the phthalates, but for other anti-androgens (e.g., androgen receptor antagonists), one might select a different effect from the suite of fetal malformations caused by these compounds. He asked

whether one could use one endpoint for one chemical and another endpoint for a second? Dr. Foster noted that it cannot just be any endpoint; it must be related to the syndrome. Dr. Tyl voiced concern that if we use different PODs and endpoints for each chemical, this may be too complicated to actually achieve.

Dr. Gennings suggested two approaches. The first would be to define the most sensitive effect for each chemical, and base the POD on that for each chemical. Another way, for which she indicated her preference, would be to define the syndrome, and ask whether each animal has any of the manifestations of the syndrome. For continuous parameters such as AGD, this would mean defining "yes" as being less than a certain length. You are after the most sensitive endpoint per animal. This would require going back to raw data and establishing some definitions. Dr. Tyl noted that this would take some statistical work to figure out what the unit is, since the dam was actually the one being dosed, while the pups are being examined. Dr. Gennings suggested that the endpoint could be the percent of pups affected.

Dr. Foster then asked panelists and discussants to discuss the strengths and weaknesses of using the HI as a method of measuring cumulative risk.

Dr. Fox advocated for the HI approach because you can accommodate many different types of chemicals, and the data that exist fit within the broader assumptions of that approach. However, she noted that there are considerations that should be addressed, including how well the data fit within the assumptions for the approaches.

Dr. Bob Benson, EPA Region 8, stated that the HI method is a relatively simple method, but it is difficult to rationalize the biology and decide which values will be used as inputs in the equation. He pointed out that this assessment would be very situation-specific, and would require disciplined judgment on how to proceed. Ms. Teuschler explained that the HI is not a risk estimate, but rather is a risk indicator, and agreed that sound judgment would be important. For example, does an HI value of 1.2 mean there will be a health problem, or not? Sometimes these judgments are not easy to answer. Dr. Fox thought the method would be useful in making risk management decisions, but that more thought and discussion would be required. Several panelists voiced concerns regarding the best way to calculate DI (exposure) values. Dr. DeVito pointed out that while the group was focused on uncertainty in the RfDs, the exposure numbers may be much more uncertain. He recommended carrying out the assessment and finding out exactly where these uncertainties lie.

Dr. DeVito then expressed confusion over the relationship between the RPF (discussed by the previous presenter) and HI approaches. In doing an HI approach, one is adding the PODs for a syndrome with different endpoints; e.g., different endpoints for different chemicals, but that wouldn't work for an RPF approach. He noted that the panel stated that RPFs are tissue-specific, and questioned whether these two methods would contradict each other. Dr. Foster ventured that you would have different relative potencies for each endpoint, rather than just one. Dr. Gray stated that if one picks the most sensitive effect for each of two chemicals with different modes of action, it would be similar to the HI approach. Dr. DeVito asked if the RPF was a better alternative than HI or just an alternative. Dr. Gray stated that there is a big difference in approaches. In some cases, it is recommended to try both, since it is not always completely clear which is more accurate.

4.3. Point of Departure Index (PODI) Method, Dr. Chris Gennings (VCU)

Dr. Chris Gennings presented a second potential method for conducting a cumulative risk assessment, the PODI method, for combined exposures to anti-androgenic chemicals, including phthalates. Her presentation was again based on the study entitled, "Combined exposures to anti-androgenic chemicals:

steps towards cumulative risk assessment" (Kortenkamp and Faust, 2010), as well as the 2008 NAS phthalate report (NRC, 2008).

Dr. Gennings began her presentation by presenting a number of figures depicting at what point aggregation occurs in various methods. In the HI approach she previously described, safety factors are applied to individual reference values derived from animal data, and individual tolerable exposures in humans are derived. From the individual tolerable exposure values, a cumulative tolerable exposure level is subsequently established. In the PODI approach, one develops a reference value for the mixture of chemicals, and then applies one set of safety factors to that combined value to reach the tolerable exposure level in humans. The difference between the two approaches, therefore, is where the uncertainty factors are applied.

Dr. Gennings then provided an example of the identification of a POD for a mixture, as illustrated by Figure 14. For a mixture of five hypothetical chemicals, dose-response data for each chemical was fit to a threshold model, and a benchmark dose lower limit (BMDL) was derived for each chemical. When all five BMDLs were combined, an effect was observed, even if each chemical's individual BMDL was below an effect. She indicated that if the BMDLs for all five chemicals (i.e., levels that theoretically should not have much of an effect) are combined, then divided by five, the resulting POD is more conservative than simply adding the individual BMDLs. Dr. Gennings noted that this gets to the question of what is the combination effect, and indicated that simply summing up the BMDLs for each individual chemical does not demonstrate the true combination effect. For mixtures, therefore, you need to use a different approach to derive the BMDL.

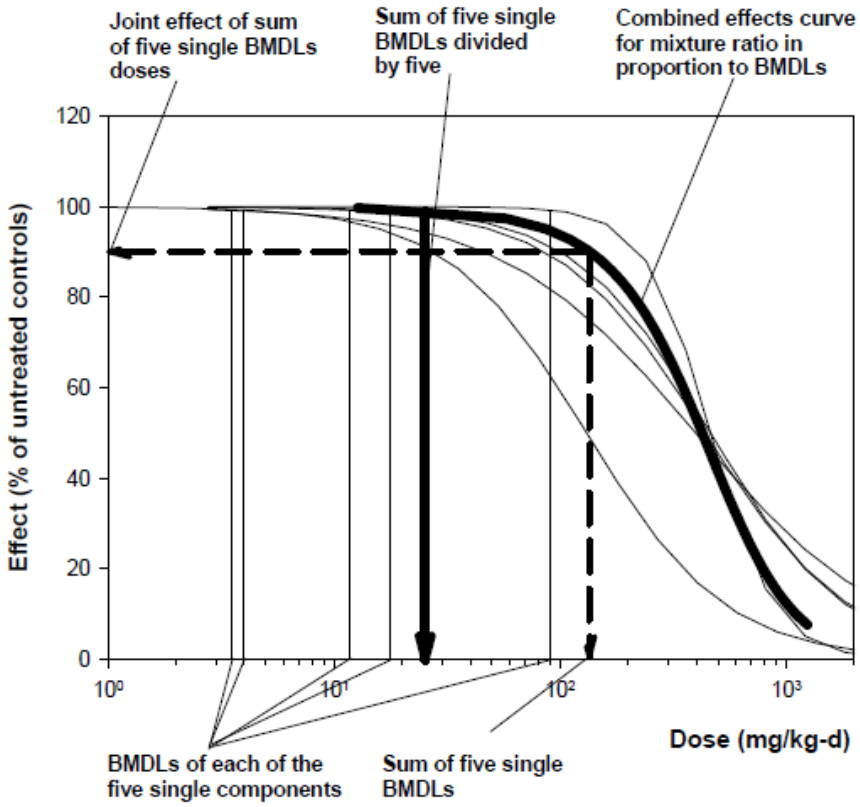

Figure 14. Schematic to illustrate the derivation of a point of departure for a mixture dose, here for a BMDL (NRC, 2008; Figure 5-5)

Dr. Gennings next explained the process of calculating the PODI. She stated that the POD for cumulative risk assessment should relate directly to relative potency of the chemicals in the group. Ideally, it should be associated with a biological response (e.g., the ED_{10}) rather than the NOAEL, which is largely an artifact of study design (e.g., the selected doses and spacing of doses). The PODI calculation is shown in Figure 15. The POD index removes uncertainty factors from each component of the sum and may use a single uncertainty factor for the mixture. It is equivalent to calculating the HI when all uncertainty factors are identical for the *c* chemicals.

$$\text{POD Index (PODI)} = \sum_{j=1}^{c} \frac{\text{DI}_j \ (\mu\text{g/kg/day})}{\text{POD}_j \ (\text{AA}; \ \mu\text{g/kg/day})}$$

Figure 15. Calculating the PODI

This method was then applied to existing data. PODs were determined based on anti-androgenicity data from Korenkamp and Faust (2010), which focused on five phthalates including DEHP, DnBP, DiBP, BBP, and DINP; other anti-androgens including vinclozolin, prochloraz, procymidone, linuron, fenitrothion, and p,p'-DDE; and chemicals with limited evidence of anti-androgenicity including BDE 99, bisphenol A, butyl paraben and propyl paraben. The PODs were based on BMDLs for the phthalates, and NOAELs for the other chemicals (Case 1) or on the assumption that DiBP, DnBP, DEHP, and BBP are approximately equipotent and the RPF for DiNP compared to DEHP is 0.15 (Case 2). After the calculations, the PODI for median chemical intake was 0.36 for Case 1 and 0.38 for Case 2. The POD fractions for vinclozolin and DEHP had a large influence on the PODI, with prochloraz, procymidone, butyl paraben, DBP, and DiBP also contributing. The remaining chemicals made up very small fractions of the total PODI. When Dr. Gennings considered high intake of these substances, the PODI was 1.97 for Case 1 and 2.12 for Case 2, with most of the driving chemicals remaining the same (with the addition of butyl paraben as a moderate contributor). Figure 16 displays a summary of the results.

POD Index (% from phthalates)	**Median intake estimates**	**High intake estimates**
POD Case 1: (Kortenkamp and Faust, 2010)	0.36 (29%)	1.97 (8%)
POD Case 2: Assuming equi-potency (Gray, 2008)	0.39 (34%)	2.12 (15%)

Figure 16. Summary results for the HI case studies.

Dr. Gennings concluded that although the results of this PODI analysis were similar to those found with hazard index methods due to similar uncertainty factors across chemicals, this is usually not the case. She also reemphasized that BMDLs should be used instead of NOAELs whenever possible.

Discussion Summary

A few of the panelists and discussants raised concerns regarding the use of the PODI approach for cumulative risk assessment. Dr. DeVito expressed concern with applying uncertainty factors after establishing a POD for a mixture, since one could not apply knowledge of kinetics in this approach. He indicated that it could be problematic if one uses default uncertainty factors for the whole class of chemicals, when real data might exist for some chemicals. Different chemicals might have very different animal-to-human extrapolations, and Dr. DeVito was concerned that the PODI approach would decrease accuracy. He suggested that BMDLs could potentially be converted to human equivalency doses (HEDs) first, and then the uncertainty factors applied afterward.

Dr. Gennings responded that she was unsure whether the EPA would use the PODI approach for the assessment of phthalates. She indicated that this methodology could be used for cumulative assessment, but it would not necessarily be used instead of the HI method. Dr. Foster clarified that the uncertainty lies in how one is going to apply the uncertainty factors and how one can adjust for information that would improve the accuracy. He noted that the potential difficulty in incorporating data could be considered a limitation of the PODI approach, since in general, researchers try to move away from the default as often as possible. While Dr. Borgert agreed that the PODI method may decrease accuracy, he suggested that operating off of a BMDL rather than a NOAEL would be beneficial.

Ms. Teuschler then stated that one would be making an assumption that there is equivalent data quality and uncertainty across the chemicals if one uses an uncertainty factor for the combined mixture POD instead of applying uncertainty factors to each component prior to combining. She asked Dr. Gennings if Figure 14 suggested that the NAS recommended this approach of adding the BMDLs and dividing by the number of chemicals. Dr. Gennings replied that it was more a schematic discussion largely pointing out a combination effect than a recommendation. A BMDL for a chemical is a good boundary level, but in a combination of chemicals, if you have all five chemicals at the BMDL, you see an effect when they are combined. Ms. Teuschler noted that it is the kind of thing one could actually do, although it would not account for differences in relative potencies.

Dr. Tyl raised the question of how to deal with chemicals in the case where two chemicals with individual dose-response levels of 2 percent are summed and the resulting combined response level is 50 percent. She was concerned that using values derived from studies of individual chemicals to derive values for combinations of chemicals would not be equivalent to a study of exposure to both chemicals. Ms. Teuschler replied that in this case, you are assuming dose addition. If one has data that says you have interactions that would lead to synergy or antagonism, i.e., that the data does not follow dose addition, one should not be using this method. Furthermore, Ms. Teuschler noted that dose additive approaches assume levels low enough where the chemicals do not have interaction. Dr. Gennings added that this discussion is meant to be an example where if you have enough chemicals at levels where they would individually not cause any effect, that together you might see an effect, because when they act in a dose-additive manner, you end up way down the dose-response curve.

Dr. Borgert commented that when thinking about a dose response curve for one chemical, the gradation of increase can be markedly different between two equally spaced doses because of the slope changes. This steep slope is how 2 percent plus 2 percent can get to 50 percent. In his opinion, Dr. Borgert indicated the fundamental question is how the dose-response curve changes at relevant human exposure levels. Dr. Tyl suggested that this issue is important; if one has a superb study with chemical "X", and a superb study with chemical "Y", when you mix them the whole analysis can fall apart due to dose additivity or mixture interactions.

Ms. Teuschler again highlighted that dose additivity, without mixture interactions leading to synergy, will not predict the types of "2 + 2 = 50" observations seen by Dr. Gray in certain ranges of the dose-response curve. Dose additive methods assume the chemicals are at low enough concentrations where there are no mixture interactions, and thus one would assume that they do not interact at typically low environmental levels. If one had research that suggested otherwise, one should adjust the assessment to account for such data, at least qualitatively if not quantitatively.

Dr. Borgert noted that there have been some surveys looking at the incidence and magnitude of synergy, and although the studies have limitations, the incidence and magnitude of biologically significant synergy is low – that doesn't mean it doesn't happen, but its frequency is low. It's actually a useful commodity in antibiotic therapy, chemotherapy, and pesticide toxicology, so it's a marketable thing that people have tried to find, but real therapeutic synergy is difficult to find.

4.4. Other Methods/Approaches for Conducting the Cumulative Assessment of Six Selected Phthalates – A Biomonitoring Based Approach, Dr. Chris Gennings (VCU)

Dr. Chris Gennings discussed work currently being conducted to evaluate the ability to incorporate human biomonitoring data in HI approaches. In a HI, the numerator reflects an exposure estimate, in this example a daily intake (DI). Dr. Gennings and her colleagues attempted to estimate the distribution of the HI from combined exposures to anti-androgenic chemicals in women of childbearing age (19-40) and in children, using data from the 2005-2006 NHANES. Their goal was to determine whether exposure could be back calculated based on spot urine data, and to compare these estimates to what we already know about exposures. Dr. Gennings stated that they focused initially on seven phthalates including DEHP, DnBP, DIBP, BBP, DINP, DiDP (di-isodecyl phthalate), and DnOP (di-*n*-octylphthalate); they later added BOA (2-benzoxazolinone), BPB (butylparaben), and PPB (propylparaben).

Dr. Gennings continued that the researchers identified RfDs for anti-androgen effects based on Korenkamp and Faust (2010) and information from Dr. Earl Gray. DI was estimated in µg/kg-bw/day based on urinary excretion levels and the molecular weight of the parent compound. Spot sample levels were corrected for creatinine excretion rates. Based on existing data, the researchers used a tabular molar ratio of metabolite excreted to the amount of parent compound taken up. Excretion levels were set for the seven phthalates based on previous studies; however, some excretion levels were unknown, so authors had to set values (e.g., 10 percent). Dr. Gennings noted that NHANES demographic data is varied (and thus not nationally representative), and is weighted for different population groups. She also noted that she did not correct for the weights for various groups; rather, she used data from 365 women without considering their representativeness to the national population.

She noted that a difficulty with using the HI approach is determining exposure estimates that are really relevant. For example, she questioned whether one uses the median, 70th percentile, or 95th percentile? She indicated that it is unlikely that there will be someone who is going to have exposure levels at the 95th percentile level for all the chemicals. Although the reference dose (the denominator) is the same for everyone, each person may have their own personal HI based on their exposure. In the case of phthalates, everyone has a different proportion of exposures across the chemicals; everyone is their own experiment. In this exercise, we can look for covariates to see if anything is correlated with higher exposures.

Dr. Gennings noted that they found in the sample of 365 women, for phthalates only, some people in the sample had elevated HIs (HI>10 at 99th percentile), but the majority had HIs below 1 (median HI <0.2). When the exposures were broken out into the individual phthalates, most women had high HIs due to DEHP. When using a different reference dose, DINP contributed more to hazard, although it was

still not as influential as DEHP. When the additional three anti-androgens were added, the 95th percentile of exposure hazard index was above 1. As an exercise, adding an additional seven chemicals where no biomonitoring data were available, but relationships to the known chemicals existed where median exposure to the seven could be estimated, the distribution shifted further; the HI for women in the 95th percentile was 2.1. Comparing across the scenarios, the 95th percentile shifted the distribution up with high DI and with additional chemicals.

Dr. Gennings then asked how good are the exposure estimates based on back-calculating biomonitoring data relative to known exposure levels? When comparing the known estimates of median and high human intake of phthalates from the literature to the back-calculated exposures for the 365 women in the NHANES sample, most compared well, but DINP estimates did not compare well at all. Dr. Gennings noted that the DINP exposure data was from a German study, and the author of the study told her that DINP uses differ widely in Europe versus the United States. Consequently, this may not be a methodological problem as much as a reflection of the differing uses between the two countries.

Dr. Gennings concluded that the exposure estimates they made were "in the ballpark." She emphasized that biomonitoring data may provide relevant co-exposure information for cumulative risk assessments which will allow researchers to see how exposures change across a population. She noted, however, that this approach still needs to be validated.

Discussion Summary

The panelists and discussants asked Dr. Gennings clarifying questions and her opinion on the strengths and weaknesses of using this biomonitoring based approach for conducting a cumulative assessment of phthalates.

Dr. DeVito asked Dr. Gennings if her approach assumed steady state conditions, or that the exposure occurred the day before. Dr. Gennings clarified that it was most likely an assumption of steady state. In 2005–2006, NHANES took morning, afternoon, or evening urine measurements. The morning groups fasted. Exposures could have occurred throughout the day, but there was probably some assumption of steady state. This technique, she said, may under- or overestimate exposure. Dr. DeVito did not think one could overestimate exposure using this method, but he did think one could underestimate exposure.

Dr. Foster also asked if Dr. Gennings looked at how many women (if any) had all seven chemicals. Dr. Gennings replied that most women had a majority of the chemicals, indicating co-exposure to probably all of them. Only a few women had only one or two of the chemicals of interest.

Dr. Foster then asked Dr. Gennings to discuss strengths and weaknesses of this kind of approach. Dr. Gennings replied that with phthalates, fetal development is likely affected. Thus, even though the women in this study were not pregnant, they could be; thus, the data are relevant to this population. Some of the variable behavior patterns of women are probably also picked up in the NHANES data, which is useful. She noted that the main point is that if the estimates are reasonable, these exposures are real.

Ms. Teuschler commented that it was valuable that Dr. Gennings' team looked at other chemicals besides phthalates and that this "background" exposure could be used for HIs. Dr. Gennings noted that there are limitations to what chemicals are available in NHANES and right now these measurements are only for phthalates in urine, and not for other anti-androgens. Chemicals that are lipophilic would be harder to capture. Finally, Dr. Gray pointed out that several of the pesticides that these authors included have a common metabolite, so it may be hard to say which chemical was the primary exposure.

4.5. Presentation of a Case Study: Use of Phthalate Mixtures Data, Dr. Cynthia Rider (NIEHS/NTP)

Dr. Hotchkiss introduced Dr. Rider, and reviewed how the 2008 NAS phthalates committee recommended inclusion of other anti-androgens, in addition to phthalates, and the possibility that exclusion could lead to an underestimation of cumulative risk.

Dr. Rider began her presentation by reviewing mixtures theory and the specific methodologies and implications of dose addition and response addition. She noted that the classic theory behind dose addition is that there are shared modes or mechanisms (e.g., chemicals A and B bind to the same receptor), while response addition is classically applied to chemicals with different modes or mechanisms of action. Dose addition implies that you are "getting something from nothing;" in other words, chemicals present in a mixture below their individual NOAELs will contribute to the total mixture dose which can then elicit a significant effect. On the other hand, an infinite number of chemicals present in a mixture below their NOAELs will not contribute to significant effects via response addition. The key question is: when do chemicals abide by dose addition (and how similar do they have to be)?

Dr. Rider posed a question regarding the suitability of dose addition if the chemicals: have the same active metabolite, the same biochemical pathway of toxicity, the same target signaling pathway, the same target tissue, the same target organ system? Or, is it none of the above (e.g., total tumors)? She noted that, as explained in previous presentations, there is a higher degree of confidence in models for chemicals with the same active metabolite than those with less detailed information. Even though researchers do not know exactly what causes lowered fetal production of testosterone with anti-androgen exposure, their effects can be added. EPA's Office of Pesticides Programs requires that chemicals have the same biological pathway of toxicity for dose addition. In this discussion, she noted, that we will look at cases where chemicals have the same target signaling pathway or the same target tissue.

For chemicals like DBP and procymidone, they share some of the signaling pathways and abnormal effects, but not others. Dr. Rider suggested that these chemicals would only be dose additive for those endpoints for which they both intersect.

Dr. Rider then explained some of her group's mixtures modeling. In one seven-chemical mixture study in rats (3 phthalates: DBP, BBP, and DEHP; 2 androgen-receptor antagonists: vinclozolin and procymidone; and 2 mixed-mechanism compounds with the ability to lower androgen levels: prochloraz and linuron), the goal was to compare dose addition, response addition, and integrated addition (dose addition within groups, and response addition between groups). It was assumed that these chemicals had the same target signaling pathway. Dr. Rider and her colleagues found that each of the chemicals had fairly linear dose-response curves for AGD. However, for the other responses (retained nipples, hypospadias, epididymal agenesis, gubernacular agenesis, and undescended testes), the seven chemicals had much different dose-response curves. For AGD, all models yielded similar predictions of effect. However, when looking at more threshold-like effects, dose addition was much closer to the observed effects compared to the other models. When Rider et al. (2009) added two more phthalates, making a 10-chemical mixture, dose addition was the best fit compared to response addition or integrative addition. However, in a four-chemical mixture (Christiansen et al., 2009), none of the models predicted results well.

In order to test a group of chemicals with the same assumed target tissues, Dr. Rider and her colleagues combined DBP and dioxin, which both affect reproductive tissues without direct interaction with the androgen receptor. However, these two chemicals have little in common in terms of effects other than epididymal agenesis and accompanying lesions. Response addition methods resulted in lower-than-

predicted values for epididymal sperm production and other reproductive effects. When one looks at effects where they do not intersect (e.g., where only DBP produces effects), the prediction of the mixture is driven by the active chemical and the model is better able to predict it.

Dr. Rider spent the last portion of her presentation highlighting important points to consider. She noted that the hypotheses of mixture effects are specific to each target endpoint or tissue. Endpoints with linear dose-response curves (e.g., AGD) do not have different predictions resultant from dose and response addition methods. However, although response addition sometimes provides an equally good prediction compared to dose addition, it does not perform better than dose addition (and often performs poorly). Finally, chemicals with the same active metabolite, mode or mechanism of action, or signaling pathway conform to a dose addition model.

Discussion Summary

The panelists and discussants addressed some points of clarification and then renewed the debate on the applicability of dose addition versus response addition.

In terms of clarification, Ms. Teuschler asked Dr. Rider what dose levels were used in her studies. Dr. Rider replied that in the binary studies, doses were half of the ED_{50}, and around the NOAEL. For the seven-chemical mixture, dose levels were present below their NOAELs for malformations (but not AGD).

A number of panelists and discussants opined about the usefulness of dose addition over response addition. Dr. Borgert pointed out that another possible explanation for the data seen by Dr. Rider's team is that response addition fails at inflection points in the dose-response curve, but not above or below these points. He was not convinced that dose addition is always a better model at a different response range than what was used. Dr. Boekelheide commented that the certainty of the applicability of dose addition is good for similar active metabolites, biochemical pathways of toxicity, or target signal pathways, but less certain for target tissue effects. Dr. Gray explained that in fifty experiments, none showed that response addition was better than dose addition, although in some cases the models both predicted the same results. He stated that dose addition is generally a more accurate predictor of actual mixture responses observed, especially for this class of chemicals. He felt that the objective should be to find out at what point does dose addition fail to be the optimal method, e.g., fetal testicular toxicants that act at different critical periods, response addition may be more appropriate.

4.6. Summary of Day 2

Following the conclusion of the presentations, Dr. Foster asked the panelists and discussants if they had any other points or issues to discuss that had not already been covered.

Ms. Teuschler reiterated that cumulative HI methods can be used to sum across chemicals and exposure pathways, but that this topic was largely not discussed. Dr. DeVito agreed this would be applicable and that knowing the different pathways and how they contribute to overall risk would allow one to make strategic risk management decisions to reduce exposure. Ms. Teuschler also explained that it is EPA's goal to get the individual phthalate chemicals' hazard and dose-response data into the IRIS database, in addition to the combination effects.

Finally, based on the presentations and discussions throughout the day, Drs. Lambert and Hotchkiss identified and presented the following summary points. Similar to Day 1, since the goal of this workshop was to collect individual input and opinions rather than achieve group consensus, the panelists and discussants did not discuss these points in detail; rather, they reviewed the list in its entirety to ensure that additional points should not be included.

Summary Points for Day 2:

- Dose addition better fits available data than response addition. Dose addition approaches should be performed where possible.
- It may be worthwhile to compare a whole mixtures approach with a component-based approach.
- It might be useful to illustrate the PODs and RfVs for the most sensitive outcome by phthalate (from among all endpoints observed), and, identify multiple PODs and RfVs for different endpoints within an outcome (e.g., phthalate syndrome).
- Fetal testosterone levels may be a suitable endpoint for phthalates, but it may not be applicable to other anti-androgens that act on same/similar outcomes via some other mode of action (e.g., receptor antagonism).
- The critical endpoint for a given phthalate may not necessarily be the same as the endpoint(s) selected for grouping for cumulative assessment purposes.
- Cumulative hazard index could be a valuable approach as part of a cumulative assessment of phthalates due to its flexibility.
- In the future, other phthalates and other anti-androgens that affect the phthalate syndrome must be considered further.
- Human biomonitoring data may provide relevant exposure information for cumulative risk assessments, but there are many uncertainties, specifically regarding the applicability of spot urine data for phthalates and other chemicals with short half lives.

5. Summary of Public Comments

Workshop participants were invited to deliver oral comments on both days of the workshop. Three individuals provided comment on behalf of the American Chemistry Council (ACC), People for the Ethical Treatment of Animals (PETA), and Natural Resources Defense Council (NRDC).

American Chemistry Council

On Day 1, Mr. Steve Risotto spoke on behalf of the ACC Phthalate Esters panel. He stated that the panel supports EPA's proposal to focus the workshop and associated IRIS on the existing phthalates data set but warned that data were insufficient to implement the NAS's recommendation to include other non-phthalate chemicals with common adverse outcomes into the their quantitative assessment. Any attempt to incorporate other substances into this current assessment, he said, would generate little value to risk managers.

Mr. Risotto then questioned how the EPA intends to conduct a cumulative risk assessment without considering exposure. He cited the example of DPP, which has been included in mixture studies but is not a commercial product, nor is there any intent of it becoming one. He advised that any attempt to evaluate potential health effects from multiple phthalates be grounded by the practical consideration of those substances to which the public will actually be exposed.

Mr. Risotto also encouraged EPA to focus their analysis on health endpoints that are clearly adverse. He cited the example of male fetal rat testosterone, which has been called the most sensitive indicator of male rat development effects and an appropriate endpoint for risk assessment of phthalates. He questioned the usefulness of this indicator, stating that it was unclear whether all phthalates caused fetal testosterone reduction, whether such reduction results in adverse health effects collectively termed "phthalate syndrome," and whether these rat data are relevant to humans. He concluded his remarks by requesting that the EPA be fully transparent in their logic for choosing a phthalate endpoint, so that others may analyze the data on which EPA's conclusions are based.

People for the Ethical Treatment of Animals

Mr. Joseph Manuppello of PETA summarized points related to animal studies that were made at the NTP's Board of Scientific Councilors meeting during the previous week. He reviewed that:

- Research programs should address clearly testable hypotheses and developmental experiments with animals should be well planned.
- A tiered approach should be used, beginning with mathematical modeling followed by short term *in vitro* screening assays and simple experiments using limited numbers of animals.
- Robust exposure data is needed.
- Dose versus response addition questions should be addressed with simple experiments using few animals.
- Cellular responses should be evaluated definitively before moving on to animals.
- Methods employed in Tox 21, a collaboration between EPA, NIEHS/National Toxicology Program, National Institutes of Health/National Human Genome Research Institute , NIH Chemical Genomics Center, and the Food and Drug Administration to research, develop, validate and translate innovative chemical testing methods that characterize toxicity pathways, offer an opportunity to screen large numbers of environmental mixtures, and the information gained from these studies will help reduce the number of animal studies required.

National Resources Defense Council

Dr. Sarah Janssen spoke on behalf of the NRDC. While she acknowledged the progress being made regarding the cumulative risk assessment of phthalates, she expressed concern that more phthalates were not being included in EPAs IRIS assessment. She also expressed concern that EPA was not fully incorporating the recommendations from the NAS report, since the current assessment is limited to phthalates that have the same mechanism or mode of action. She encouraged EPA to include other anti-androgenic phthalates.

She encouraged EPA to incorporate information and recommendations from other NAS reports, such as "Toxicity Testing in the 21st Century" (NRC, 2007) and "Science and Decisions: Advancing Risk Assessment" (NRC, 2009). Specifically, she encouraged EPA to consider the range of human variability and susceptibility, as well as background levels of exposure in regards to phthalates.

She concluded by announcing a new clinical study she is involved with that will track phthalate exposure throughout pregnancy and record AGD for 1200 infants at birth and 12 months. She noted that while this data will not be available for incorporation into EPA's IRIS assessment, it will likely be published in the next few years.

6. References

Barlow, N.J., McIntyre, B.S., Foster, P.M. (2004). Male reproductive tract lesions at 6, 12, and 18 months of age following in utero exposure to di(n-butyl) phthalate. *Toxicologic Pathology* 32: 79-90.

Berenbaum, M.C. (1989). What is synergy? *Pharmacological* Reviews 41: 93-141.

Blystone, C.R., Kissling, G.E., Bishop, J.B., Chapin, R.E., Wolfe, G.W., Foster, P.M. (2010). Determination of the di-(2-ethylhexyl) phthalate NOAEL for reproductive development in the rat: importance of the retention of extra animals to adulthood. *Toxicological Sciences* 116: 640-6.

Christiansen, S., Scholze, M., Dalgaard, M., Vinggaard, A.M., Axelstad, M., Kortenkamp, A., Hass, U. (2009). Synergistic disruption of external male sex organ development by a mixture of four antiandrogens. *Environmental Health Perspectives* 117: 1839-46.

Gray, L.E., Barlow, N.J., Howdeshell, K.L., Ostby, J.S., Furr, J.R., Gray, C.L. (2009). Transgenerational effects of di (2-ethylhexyl) phthalates in the male CRL: CD(SD) rat: Added value of assessing multiple offspring per litter. *Toxicological Sciences* 110: 411-25.

Hauser, R., Meeker, J., Duty, S., Silva, M., Calafat, A. (2006). Altered semen quality in relation to urinary concentrations of phthalate monoester and oxidative metabolites. *Epidemiology* 12: 682-91.

Hotchkiss, A.K., Parks-Saldutti, L.G., Ostby, J.S., Lambright, C., Furr, J., Vandenbergh, J.G., Gray, L.E. Jr. (2004). A mixture of the "antiandrogens" linuron and butyl benzyl phthalate alters sexual differentiation of the male rat in a cumulative fashion. *Biology of Reproduction* 71: 1852-61.

Howdeshell, K.L., Wilson, V.S., Furr, J., Lambright, C.R., Rider, C.V., Hotchkiss, A.K., Gray, L.E. Jr. (2008). A mixture of five phthalate esters inhibits fetal testicular testosterone production in the Sprague-dawley rat in a cumulative, dose-additive manner. *Toxicological Sciences* 105: 153-65.

Huang, P. , Kuo, P., Chou, Y., Lin, S., Lee, C. (2009). Association between prenatal exposure to phthalates and the health of newborns. *Environmental International* 35: 14-20.

Jönsson, B.A.G., Richthoff, J., Rylander, L., Giwercman, A., Hagmar, L. (2005). Urinary phthalate metabolites and biomarkers of reproductive function in young men. *Epidemiology* 16: 487-93.

Kortenkamp, A. and Faust, M. (2010). Combined exposures to anti-androgenic chemicals: steps towards cumulative risk assessment. *International Journal of Andrology* 33: 463-74.

Meeker, J., Calafat, M., Hauser, R. (2009). Urinary Metabolites of Di(2-ethylhexyl) Phthalate are associated with decreased steroid hormone levels in adult men. *Journal of Andrology* 30: 287-97.

National Research Council (NRC). 2008. Phthalates and Cumulative Risk Assessment: The Tasks Ahead. The National Academies Press, Washington, DC [online]. Available at: http://www.nap.edu/catalog.php?record_id=12528 [last accessed on January 14, 2011].

Ormond, G., Nieuwenhuijsen, M.J., Nelson, P., Toledano, M.B., Iszatt, N., Geneletti, S., Elliott, P. (2009). Endocrine disruptors in the workplace, hair spray, folate supplementation, and risk of hypospadias: case-control study. *Environmental Health Perspectives* 117: 303-7.

Pan, G., Hanaoka, T., Yoshimura, M., Zhang, S., Wang, P., Tsukino, H., Inoue, K., Nakazawa, H., Tsugane, S., Takahashi, K. (2006). Decreased serum free testosterone in workers exposed to high levels of di-n-butyl phthalate (DBP) and di-2-ethylhexyl phthalate (DEHP): A cross-sectional study in China. *Environmental Health Perspectives* 114: 1643-8.

Rider, C.V., Wilson, V.S., Howdeshell, K.L., Hotchkiss, A.K., Furr, J.R., Lambright, C.R., Gray, L.E. Jr. (2009). Cumulative effects of in utero administration of mixtures of "antiandrogens" on male rat reproductive development. *Toxicologic Pathology* 37: 100-13.

Salazar-Martinez, E., Romano-Riquer, P., Yanez-Marquez, E., Longnecker, M., Hernandez-Avila, M. (2004). Anogenital distance in human male and female newborns: a descriptive, cross-sectional study. *Environmental Health;* 3: 8. doi:10.1186/1476-069X-3-8.

Sharpe RM. (2009), Male reproductive health disorders and the potential role of exposure to environmental chemicals. *Report for ChemTrust* [online]. Available at: http://www.chemtrust.org.uk/documents/ProfRSHARPE-MaleReproductiveHealth-CHEMTrust09.pdf [last accessed on January 14, 2011].

Swan, S. (2008). Environmental phthalate exposure in relation to reproductive outcomes and other health endpoints in humans. *Environmental Health* 108: 177-184.

Swan, S.H., Main, K.M., Stewart, S.L., Kruse, R.L., Calafat, A.M., Mao, C.S. Redmon, J.B., Ternand, C.L., Sullivan, S. Teague, J.L. (2005). Decrease in anogenital distance among male infants with prenatal phthalate exposure. *Environmental Health Perspectives* 113: 1056-61.

U.S. Environmental Protection Agency (EPA). 2000. Supplementary Guidance for Conducting Health Risk Assessment of Chemical Mixtures. EPA/630/R-00/002. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. August 2000 [online]. Available at: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=20533 [last accessed January 14, 2011].

Vrijheid, M., Armstrong, B., Dolk, H., van Tongeren, M., Botting, B. (2003). Risk of hypospadias in relation to maternal occupational exposure to potential endocrine disrupting chemicals. *Occupational and Environmental Medicine* 60: 543-50.

Wolff, M.S., Teitelbaum, S.L., Pinney, S.M., Windham, G., Liao, L., Biro, F., Kushi, L.H., Erdmann, C., Hiatt, R.A., Rybak, M.E., Calafat, A.M. (2010). Investigation of relationships between urinary biomarkers of phytoestrogens, phthalates, and phenols and pubertal stages in girls. *Environmental Health Perspectives* 118: 1039-46.

Appendix A. Final Agenda

Day 1: 8:00 am to 5:00 pm on December 8, 2010

8:00-8:30 Welcome and Overview
- Kevin Teichman, DAA, ORD, EPA

8:30-8:45 Workshop Expectations and Introduction of Individual Expert Panelists and Discussants
- Presenters: Andrew Hotchkiss, Jason Lambert, Stan Barone (EPA/ORD/NCEA)
    - Meeting Facilitator: Paul Foster (NIEHS/NTP)

8:45-9:45 Public comment period

9:45-10:00 Break

10:00-10:30 Overview of the NAS Recommendations on Cumulative Risk Assessment for the Phthalates
- Intro: Stan Barone (EPA/ORD/NCEA)
- Presenter: Paul Foster (NIEHS/NTP)

10:30-11:00 Overview of the Six Selected Phthalates and the Phthalate Syndrome
- Presenter: Andrew Hotchkiss (EPA/ORD/NCEA)

11:00-12:15 Human Data for Phthalates
- Intro: Jason Lambert (EPA/ORD/NCEA)
- Presenter: Glinda Cooper (EPA/ORD/NCEA)
- Discussion with Expert Panelists and Discussants

12:15-1:30 Lunch (on your own)

1:30-3:15 Animal Data (including mixtures data) for Phthalates:
Male Developmental/ Reproductive Toxicity
- Intro: Andrew Hotchkiss (EPA/ORD/NCEA)
- Presenter: Earl Gray (EPA/ORD/NHEERL)
- Discussion with Expert Panel and Discussants

3:15-3:30 Break

3:30-4:15 Animal Data for Phthalates:
Other Toxicities Associated with Phthalates
- Presenter: Jason Lambert (EPA/ORD/NCEA)
- Discussion with Expert Panel and Discussants

4:15-5:00 Wrap-up/Summary of Day 1

Day 2: 8:00 am to 5:00 pm on December 9, 2010

8:00-8:15 Welcome -Rebecca Clark, CD, NCEA, ORD, EPA

8:15-8:30 Workshop Expectations

- Presenters: Andrew Hotchkiss, Jason Lambert (EPA/ORD/NCEA)

8:30-8:45 Recap of Day 1-Paul Foster (NIEHS/NTP)

8:45-10:00 Methods for Conducting the Cumulative Assessment of Six Selected Phthalates-Hazard Index Method (NAS recommendation)

- Intro: Jason Lambert (EPA/ORD/NCEA)
- Presenter: Chris Gennings (Virginia Commonwealth Univ)
- Discussion with Expert Panel and Discussants

10:00-10:15 Break

10:15-11:15 Methods for Conducting the Cumulative Assessment of Six Selected Phthalates-Point of Departure Index Method (NAS recommendation)

- Intro: Jason Lambert (EPA/ORD/NCEA)
- Presenter: Chris Gennings (Virginia Commonwealth Univ)
- Discussion with Expert Panel and Discussants

11:15-12:30 Lunch (on your own)

12:30-2:15 Methods for Conducting the Cumulative Assessment of Six Selected Phthalates-Other Methods/Approaches (e.g., Relative Potencies)

- Intro: Jason Lambert (EPA/ORD/NCEA)
- Presenters: Chris Gennings (Virginia Commonwealth Univ) and Linda Teuschler (EPA/ORD/NCEA)
- Discussion with Expert Panel and Discussants

2:15-2:30 Break

2:30-3:15 Methods for Conducting the Cumulative Assessment of Six Selected Phthalates-Other Methods/Approaches (Continued)

3:15-4:15 Presentation of a Case Study: Use of Phthalate Mixtures Data (phthalate + pesticide)

- Intro: Andrew Hotchkiss (EPA/ORD/NCEA)
- Presenter: Cynthia Rider (NIEHS/NTP)
- Discussion with Expert Panel and Discussants

4:15-5:00 Wrap-up/Summary of Day 2

Appendix B. List of Facilitator, Panelists, and Discussants

Meeting Facilitator

Paul Foster
NIH/NIEHS
Foster2@niehs.nih.gov

List of Panelists

Kim Boekelheide
Brown University
Kim_Boekelheide@Brown.edu

Mike DeVito
NIH/NIEHS
devitom@niehs.nih.gov

Mary Fox
Johns Hopkins University
mfox@jhsph.edu

Chris Gennings
Virginia Commonwealth University
gennings@vcu.edu

Earl Gray
EPA/ORD
Gray.earl@epa.gov

Paul Lioy
University of Medicine and Dentistry of New Jersey
plioy@eohsi.rutgers.edu

Barry McIntyre
NIH/NIEHS
mcintyrebs@niehs.nih.gov

Linda Teuschler
EPA
Teuschler.linda@epa.gov

List of Discussants

Katherine Anitole
EPA/OPPT
anitole.katherine@epa.gov

Norman Barlow
Sanofi-Aventis
Norman.barlow@sanofi-aventis.com

Bob Benson
EPA Region 8
Benson.bob@epa.gov

Chris Borgert
Applied Pharmaceutical Toxicology, Inc.
cjborgert@apt-pharmatox.com

Kembra Howdeshell
NIH/NIEHS
howdeshellkl@niehs.nih.gov

Cynthia Rider
NIH/NIEHS
ridercv@niehs.nih.gov

Rochelle Tyl
Research triangle Institute
rwt@rti.org

Vickie Wilson
EPA
Wilson.vickie@epa.gov

Appendix C. Workshop Registrants

Last Name	First Name	Affiliation	E mail
Alger	Heather	Pew Charitable Trusts	halger@pewtrusts.org
Anitole	Katherine	EPA/OPPT	anitole.katherine@epa.gov
Arzuaga	Xabier	EPA/ORD	arzuaga.xabier@epa.gov
Babich	Michael	U.S. Consumer Product Safety Commission (CPSC)	mbabich@cpsc.gov
Bale	Ambuja	EPA/ORD	bale.ambuja@epa.gov
Barlow	Norman	Sanofi-Aventis	Norman.barlow@sanofi-aventis.com
Barone	Stan	EPA/ORD	barone.stan@epa.gov
Beck	Nancy	Office of Management and Budget	nbeck@omb.eop.gov
Benson	Bob	EPA Region 8	benson.bob@epa.gov
Birchfield	Norm	EPA/ORD	Birchfield.Norman@epamail.epa.gov
Boekelheide	Kim	Brown University	Kim_Boekelheide@Brown.edu
Borgert	Chris	Applied Pharmaceutical Toxicology, Inc.	cjborgert@apt-pharmatox.com
Bottimore	David	Versar Inc.	dbottimore@versar.com
Brown	Ron	FDA	ronald.brown@fda.hhs.gov
Brown	Paul	National Research Center for Women and Families	pb@center4research.org
Butala	JH	Ferro, Inc.	butala@jhbutala.com
Cai	Christine	EPA/ORD	cai.christine@epa.gov
Callahan	Michael	MDB, Inc.	mcallahan@michaeldbaker.com
Carlson	Kent	CPSC	kcarlson@cpsc.gov

Last Name	First Name	Affiliation	E mail
Christensen	Krista	EPA	christensen.krista@epa.gov
Cinalli	Christina	EPA/OPPT	cinalli.christina@epa.gov
Claassen	Ann	Latham & Watkins	ann.claassen@lw.com
Cooper	Glinda	EPA/ORD	cooper.glinda@epa.gov
Deener	Kacee	EPA/ORD	deener.kathleen@epa.gov
DeVito	Mike	NIH/NIEHS	devitom@niehs.nih.gov
Deyo	James	Eastman Chemical Company	deyo@eastman.com
Differding	Kelly	ICF International	KDifferding@icfi.com
Doyle	Elizabeth	EPA/OW	doyle.elizabeth@epa.gov
Euling	Susan	EPA/ORD	Euling.Susan@epamail.epa.gov
Flowers	Lynn	EPA/ORD	flowers.lynn@epa.gov
Foster	Paul	NIH/NIEHS	foster2@niehs.nih.gov
Fox	Mary	Johns Hopkins University	mfox@jhsph.edu
Francis	Elaine	EPA/ORD	francis.elaine@epa.gov
Galizia	Audrey	EPA/ORD	galizia.audrey@epa.gov
Gennings	Chris	VCU	gennings@vcu.edu
Gimlin	Peter	EPA/OPPT	gimlin.peter@epa.gov
Gordon	Ami	ICF International	AGordon3@icfi.com
Graf	Ryan	Policy Navigation Group	rgraf@policynavigation.com
Gray	Earl	EPA/ORD	gray.earl@epa.gov
Guyton	Kate	EPA/ORD	guyton.kate@epa.gov

Last Name	First Name	Affiliation	E mail
Hegstad	Maria	Risk Policy Report	mhegstad@iwpnews.com
Hogan	Karen	EPA/ORD	hogan.karen@epa.gov
Hotchkiss	Andrew	EPA/ORD	hotchkiss.andrew@epa.gov
Howdeshell	Kembra	NIEHS/NTP	howdeshellkl@niehs.nih.gov
Janssen	Sarah	National Resources Defense Council	sjanssen@nrdc.org
Jinot	Jennifer	EPA/ORD	jinot.jennifer@epa.gov
Jones	Samantha	EPA/ORD	jones.samantha@epa.gov
Kapal	Dewan	FDA	kepal.dewan@fda.hhs.gov
Kransler	Kevin	ExxonMobil Biomedical Sciences	kevin.kransler@exxonmobil.com
Lambert	Jason	EPA/ORD	lambert.jason@epa.gov
Lioy	Paul	UMDNJ	plioy@eohsi.rutgers.edu
Lorber	Matt	EPA/ORD	Lorber.Matthew@epamail.epa.gov
Lowit	Anna	EPA/OPP	Lowit.Anna@epamail.epa.gov
Makris	Sue	EPA/ORD	makris.susan@epa.gov
Manuppello	Joseph	PETA	josephm@peta.org
McIntyre	Barry	NIH/NIEHS	mcintyrebs@niehs.nih.gov
Miller	Mark	EPA/OEI	Miller.Mark@epamail.epa.gov
Newhouse	Kathleen	EPA/ORD	newhouse.kathleen@epa.gov
Patton	Geoffrey	FDA	geoffrey.patton@fda.hhs.gov
Petersen	Dan	EPA/ORD	petersen.dan@epa.gov

Last Name	First Name	Affiliation	E mail
Poster	Dianne	National Institute of Standards and Technology	poster@nist.gov
Pratt	Margaret	EPA/ORD	pratt.margaret@epa.gov
Putzrath	Resha	Navy and Marine Corps Public Health Center	resha.Putzrath@med.navy.mil
Raffaele	Kathleen	EPA/OPP	raffaele.kathleen@epa.gov
Rider	Cynthia	NIH/NIEHS	ridercv@niehs.nih.gov
Rieth	Susan	EPA/ORD	rieth.susan@epa.gov
Risotto	Steve	ACC	steve_risotto@americanchemistry.com
Rollins	Jim	Policy Navigation Group	jerollins@policynavigation.com
Rutigliano	Marian	EPA/ORD	rutigliano.marian@epa.gov
Salmon	Andy	CalEPA	ASALMON@oehha.ca.gov
Sasso	Alan	EPA	sasso.alan@epa.gov
Scott	Cheryl	EPA/ ORD	scott.cheryl@epa.gov
Segal	Deb	EPA/ORD	Segal.Deborah@epamail.epa.gov
Sheth	Christopher	EPA/ORD	sheth.christopher@epa.gov
Simpson	Heather	ICF International	HSimpson@icfi.com
Sonanwane	Bob	EPA/ORD	sonawane.bob@epa.gov
Strong	Jamie	EPA/ORD	jstrong@ceq.eop.gov
Teuschler	Linda	EPA/ORD	teuschler.linda@epa.gov
Tuxen	Linda	EPA/ORD	tuxen.linda@epa.gov
Tyl	Rochelle	RTI	rwt@rti.org
Vulimiri	Sury	EPA/ORD	vulimiri.sury@epa.gov

Last Name	First Name	Affiliation	E mail
Walker	Teneille	EPA/ORD	walker.teneille@epa.gov
Wilson	Vickie	EPA/ORD	wilson.vickie@epa.gov
Zachek	Christine	EPA/OCHP	zachek.christine@epa.gov
Zuckerman	Diana	National Research Center for Women and Families	dz@center4reseacrch.org