TERA

INDEPENDENT
NON-PROFIT
SCIENCE
FOR PUBLIC HEALTH
PROTECTION

Peer Review of the CHAP Draft Report on Phthalates and Phthalate Substances

Submitted to: Consumer Product Safety Commission (CPSC)

FINAL, August 12, 2013

Submitted by: Toxicology Excellence for Risk Assessment

Contact: Jacqueline Patterson

(patterson@tera.org)

2300 Montana Avenue, Suite 409

Cincinnati, Ohio 45211

t. 513-542-RISK (7475) x 29

www.TERA.org

This page left intentionally blank.

NOTE

This report was prepared by scientists of Toxicology Excellence for Risk Assessment (TERA). The peer reviewers served as individuals, representing their own personal scientific opinions. They did not represent their companies, agencies, funding organizations, or other entities with which they are associated. Their opinions should not be construed to represent the opinions of their employers or those with whom they are affiliated.

This page left intentionally blank.

TABLE OF CONTENTS

This page left intentionally blank.

1 Introduction and Background

Toxicology Excellence for Risk Assessment (TERA) arranged for written peer review of the draft Consumer Product Safety Commission's (CPSC) report entitled *Chronic Hazard Advisory Panel on Phthalates and Phthalate Alternatives (May 15, 2013).* The goal of the expert review was to provide CPSC and the Chronic Hazard Advisory Panel (CHAP) with independent scientific and technical expert opinion and comment on the draft text. The objective of the peer review was to obtain a broad, high-level peer review of the report, focusing on the overall risk assessment process that the CHAP applied to phthalates, and in particular on the novel methods the CHAP used (e.g., development of distributions of hazard indices for cumulative risk). The experts provided their own personal opinions, and did not represent the opinions of their employers or other organizations with whom they may be affiliated. The information in this report does not represent the opinions of Toxicology Excellence for Risk Assessment. TERA performed this work for CPSC under contract CPSC-D-12-0001.

The following is a list of phthalates and phthalate substitutes used in this report, with CAS numbers and abbreviations noted for each.

Chemical Name	Abbreviation	CAS Number
Acetyl Tributyl Citrate	ATBC	77-90-7
Butylbenzyl Phthalate	BBP	85-68-7
Dicyclohexyl Phthalate	DCHP	84-61-7
Di(2-ethylhexyl) adipate	DEHA	103-23-1
Di(2-ethylhexyl) Phthalate	DEHP	117-81-7
Di(2-ethylhexyl) terephthalate	DEHT	6422-86-2
Di(2-propylheptyl) Phthalate	DPHP	53306-54-0
Diethyl Phthalate	DEP	84-66-2
Diisobutyl Phthalate	DIBP	84-69-5
Diisodecyl Phthalate	DIDP	26761-40-0 and 68515-49-1
Diisononyl hexahydrophthalate	DINX	166412-78-8
Diisononyl Phthalate	DINP	28553-12-0 and 68515-48-0
Diisooctyl Phthalate	DIOP	27554-26-3
Dimethyl Phthalate	DMP	131-11-3
Di-*n*-butyl Phthalate	DBP	84-74-2
Di-*n*-hexyl Phthalate	DHEXP	84-75-3
Di-*n*-octyl Phthalate	DNOP	117-84-0
Di-*n*-pentyl Phthalate	DPENP	131-18-0
2,2,4-Trimethyl-1,3 pentanediol diisobutyrate	TPIB	6846-50-0
Tris(2-ethylhexyl) trimellitate	TOTM	3319-31-1

1.1 Conflict of Interest and Non-Disclosure

TERA conducted a search and evaluation of current and completed TERA projects to determine whether TERA would have any conflicts of interest in organizing and conducting the peer review for the CHAP report. TERA did not identify any current projects that involve phthalates. TERA completed several projects involving individual phthalates over five years prior to this task, but none of that work created a conflict of interest.

TERA staff and the peer reviewers signed non-disclosure agreements that preclude them from discussing or disclosing the documents and information that was provided by CPSC, which is otherwise not publicly available or previously known to the individual.

1.2 Reviewer Selection

The CPSC determined that the scientists who would peer review the draft CHAP report should be selected using the same criteria, process, and restrictions as were used for selecting CHAP members. These criteria are described in the contract statement of work (page 2):

> "Specifically, the peer reviewers must be scientists:
> 1) who are nominated by the National Academy of Sciences;
> 2) who are not officers or employees of the United States (other than employees of the National Institutes of Health, the National Toxicology Program, or the National Center for Toxicological Research), and who do not receive compensation from or have any substantial financial interest in any manufacturer, distributor, or retailer of a consumer product; and
> 3) who have demonstrated the ability to critically assess chronic hazards and risks to human health presented by the exposure of humans to toxic substances or as demonstrated by the exposure of animals to such substances."

The President of the National Academy of Sciences nominated 20 scientists to be considered and CPSC screened the scientists for availability and the CPSC Office of the General Counsel evaluated them for conflict of interest. CPSC forwarded to TERA the names of five experts with collective experience in risk assessment, biomonitoring data, biostatistics, toxicology, and phthalates; four of these experts agreed to participate. TERA also screened the selected experts for potential conflict of interest. The following four experts reviewed the draft CHAP document and provided written comments that are captured in full in this report:

- Paul Foster, Ph.D., Chief of the Toxicology Branch, National Institute of Environmental Health Sciences, Durham, NC, USA
- Judith A. Graham, Ph.D., retired from the U.S. EPA and American Chemistry Council, part-time consulting, Pittsboro, NC USA
- Donna Vorhees, Ph.D., The Science Collaborative, Ipswich, MA, USA

- Yiliang Zhu, Ph.D., Professor, Department of Epidemiology and Biostatistics, University of South Florida, Tampa, FL, USA

1.3 Charge Development

A key aspect of a successful peer review is a comprehensive list of objective questions to frame the reviewers' comments and ensure that the reviewers are focused on the most important issues. CPSC provided a list of draft charge questions with the Statement of Work. TERA reviewed the draft charge questions and draft document and recommended several additional questions and revisions to improve clarity and objectivity of the charge. The charge questions focused on the adequacy, quality and relevance of the data and information and whether the conclusions reached were supported by the data. Both focused and open-ended questions were used to provide reviewers with the opportunity to identify and discuss all the issues they felt were important. A copy of the charge and instructions for reviewers is found in Appendix A.

1.4 Review Package and Pre-Review Teleconference

The review package TERA sent to the experts included the draft CHAP report and appendices, instructions for reviewers, the charge questions, and references. The experts were allotted approximately 6 weeks for their review. Prior to the start of the review, TERA held a teleconference with the experts and CPSC staff to provide background information on the CHAP process and document, explain the charge and review process, and answer reviewers' questions. Appendix B contains information on reviewer clarifying questions and additional information provided to the reviewers. Slides used by CPSC staff to provide background are found in Appendix C.

1.5 Reviewer Comments and TERA Report

TERA compiled the reviewers' comments by charge question, randomly assigning each reviewer a reviewer number that was used throughout the report. The assigned reviewer number is meant to keep each reviewer's specific comments anonymous, although the names and affiliations of the reviewers are provided. TERA staff screened the experts' comments for completeness and clarity, and the CPSC was given the opportunity to review the peer reviewers' comments and submit to TERA clarifying questions for the reviewers.

The experts' comments were compiled into this comprehensive report entitled *Peer Review of the CHAP Draft Report on Phthalates and Phthalate Substances.* CPSC reviewed the draft peer review report and had several clarifying questions for reviewers and the text below reflects the reviewers' final comments.

2 Peer Reviewers Responses to Charge Questions

2.1 Analysis of Biomonitoring Data

2.1.1 Question 1a: Is the CHAP's analysis of biomonitoring data appropriate for assessing cumulative risk?

2.1.1.1 Reviewer 1:

This approach seems very reasonable as this is probably a robust dataset on human exposure and the best that is likely to be available. However did the Panel look at other datasets that might not be as large, but are targeted at the most sensitive life stage? For example, have they looked also at other fluids that might be available with these lesser sized cohorts – e.g. amniotic fluid (AF) – human AF banks do exist and perhaps could be used to get a better assessment of fetal exposure closer to the critical window for induction of effects on sexual differentiation than, for example, cord blood or even mothers' blood? Amniotic fluid measurements of phthalates give a more direct measure of fetal exposure (Calafat *et al*., 2006; Silva *et al*., 2004) and show a difference between dam and fetal exposure, the degree of phase 2 metabolism – this may put the fetus at increased risk for more free monoester exposure than conjugates. There may also be a recirculation of the amniotic fluid with the fetus swallowing fluid and then excreting metabolites into the amniotic fluid. A comparison of AF levels of metabolites in animals and humans has been conducted – for animals this can be obtained at the correct developmental exposure window and thus perhaps give a better comparison of internal dose and fetal exposure. This may alter how comparisons of reference doses (RfDs) are made on an internal dose estimate, rather than external applied dose levels.

2.1.1.2 Reviewer 2:

Yes. The NRC report, Human Biomonitoring for Environmental Chemicals (2006; http://www.nap.edu/catalog.php?record_id=11700) was not in the reference list. This report has extensive discussion of the utility of human biomonitoring (HBM) data in risk assessment and includes examples with phthalates. Thus, it should be used.

2.1.1.3 Reviewer 3:

Yes. The CHAP's selection and analysis of biomonitoring data is generally appropriate for quantifying cumulative risk.

Any assessment of cumulative risk involves quantification of exposure to multiple stressors that might act together to cause an adverse effect. Human biomonitoring data are generally well-suited to this purpose because they reveal the chemical mixtures to which individuals are exposed via all sources and routes of exposure. However, the data

must be of sufficient quality and relevant to the specific questions at hand.

The CHAP quantified the exposure of pregnant women, other adults, and infants using spot urine samples from the National Health and Nutrition Examination Surveys (NHANES) and the Study for Future Families (SFF). These samples were collected between 2005-2006 and 1999-2005, respectively. The CHAP provides good justification for use of these data, which appear to be the best available for representing phthalate exposure in the U.S. general population. Some other biological matrices (e.g., blood) have been sampled and analyzed for phthalates, but not as part of a large systematic sampling program like NHANES.

The spot urine HBM data were used to estimate daily intakes (DIs) of phthalates intended to reflect exposure over a time interval relevant to the health outcome of concern. The data were not collected specifically for this purpose, so it is not surprising that they have limitations, many of which have been described well in Section 4.1.3 of the draft report. This section concludes with the statement that

> "uncertainties regarding HBM data and dose extrapolations based on HBM data are within one order of magnitude, and certain factors for the possibility of overestimation of daily intake (and therefore the Hazard Index [HI]) seem to be balanced by factors for the underestimation of the DI/HI. Human biomonitoring data therefore provides a reliable and robust measure of estimating the overall phthalate exposure and resulting risk (page 64, lines 1926-1930)."

This argument would be more compelling if sources of uncertainty were presented in a manner more conducive to quantitative review and assessment. I am not suggesting that the CHAP's task warrants a fully probabilistic assessment, but simply a more explicit tabulation of uncertainties where possible. To this end, the following table summarizes limitations (primarily those identified by the CHAP), whether they are more likely to overestimate or underestimate current daily intake (DI) of phthalate mixtures that might cause adverse male developmental effects, and whether the degree of uncertainty can be readily quantified. The CHAP's conclusion that "uncertainties regarding HBM data and dose extrapolations based on HBM data are within one order of magnitude" could be true, but this statement does not seem to add much value without acknowledgment of sources of uncertainty that cannot be readily quantified. A more complete discussion of uncertainty can lead to the identification of ways to optimize collection and analysis of HBM data for use in answering risk-based questions faced by regulators.

It might be valuable to think about sources of uncertainty relevant to specific phthalates instead of simultaneously talking about all or large groupings of phthalates. For example, DEHP exposure is believed to be dominated by dietary exposure and DEHP exposure, in turn, drives risk calculations. As a result, it would be of special interest to know the set of uncertainties unique to DEHP and how they influence the level of confidence associated with HI predictions for DEHP (e.g., DEHP is among those phthalates not so well represented by single spot urine samples, and collection of urine samples after fasting and/or in the morning instead of the evening might underestimate DEHP exposure from the diet). The table below of uncertainties associated with HBM data and their use is not necessarily complete and should be completed with information from other reviews, CPSC, and the CHAP itself.

Source of Uncertainty in HBM Data for Quantifying Current Daily Intake (DI) of Phthalate Mixtures	Does uncertainty lead more likely to overestimate of DI, underestimate of DI, or neither?	Can the uncertainty be quantified?	Source
HBM data were collected before the Consumer Product Safety Improvement Act (CPSIA) of 2008 restricted some of the uses of the five antiandrogenic phthalates for which HIs were calculated.	Overestimate. This source of uncertainty especially warrants discussion because one of the phthalates banned by the CPSIA, DEHP, dominated HI calculations in this report.	No. Quantification would require an understanding of how and when phthalate exposure changed following the ban and the pharmacokinetics of individual phthalates	CHAP report, page D-49, lines 957-960
NHANES and SFF studies not designed specifically to identify members of the population with the highest exposures	Possibly underestimate upper percentile exposures	Some exposures are not specifically addressed by large surveys such as NHANES and may lead to phthalate exposures well above those found in the general population (e.g., use of certain medications with enteric coatings containing phthalates or DEHP-containing medical devices). However, some of these exposures should no longer occur with the bans required by CPSIA 2008	Calafat and McKee 2006
Limitations of the NHANEs and SFF data sets resulted in HIs being estimated for only 5 antiandrogenic phthalates. The CHAP identifies some other potentially antiandrogenic phthalates (e.g., DPENP, DCHP).	Underestimate	Not readily, but likely to be insignificant if use and exposure is as limited as the CHAP report indicates.	CHAP report, Section 5
Variability in metabolite concentration due to analytical variability	neither	± 1.1-1.2	CHAP report, page 62, lines 1831-1841
Temporal variability of metabolite concentrations – single spot urine sample "moderately predictive of each subject's exposure over 3 months" although more predictive for low molecular weight than high molecular weight phthalates. The CHAP cites a study that found within person variability was the main contributor to total variance for a high molecular	Possible overestimate of upper percentile, but should consider more recently collected data, e.g., Frederiksen et al 2013). The CHAP argues that DI and HI calculations are population-based, allowing them to assume that the HBM accurately reflect variability of exposure in the subpopulation of interest. However, at least for higher molecular weight phthalates (e.g., DEHP), it	To some extent, yes, by consulting studies with multiple urine sampling events over time for each study participant and comparing variance using single spot sample for each versus using time-integrated average for each.	CHAP report, page 63, lines 1861-1880.

Source of Uncertainty in HBM Data for Quantifying Current Daily Intake (DI) of Phthalate Mixtures	Does uncertainty lead more likely to overestimate of DI, underestimate of DI, or neither?	Can the uncertainty be quantified?	Source
weight phthalate (MEHHP).	seems possible that the upper percentiles of exposure are somewhat overestimated because of greater variability estimated using a single urine concentration instead of a time-integrated average urine concentration for each HI calculation.		
Effect of time of day for urine sample collection on metabolite concentration	Underestimate Concentrations of some phthalates, notably food-borne phthalates, are higher in the evening than in the morning. However, biomonitoring data were not collected in the evening	-1.5 (ideally note correlation detected between fasting and time of day sample is collected; e.g., Saravanabhavan et al 2013)	CHAP report, page 63, lines 1882-1889
Interindividual variability in metabolism of phthalates	Neither	± 1.2	CHAP report, page 62, lines 1852-1854
Variability in metabolite concentration due to fasting required for NHANES	Underestimate in NHANES; not relevant to SFF	Less than a factor of -2	CHAP report, page 63, lines 1896-1906
Variability due to elimination kinetics and spot samples	Underestimate for some phthalates and overestimate for other phthalates	Possible factors of -4 for low molecular weight phthalates and +2 for high molecular weight phthalates	CHAP report, page 64, lines 1908-1918
Use of creatinine correction model	Underestimate	Factor of -2	CHAP report, page 64, lines 1920-1924

Calafat AM, McKee RH. (2006). Integrating biomonitoring exposure data into the risk assessment process: Phthalates [Diethyl Phthalate and Di(2-ethylhexyl) Phthalate] as a Case Study, Environ Health Perspect. 114(11):1783–1789.

Frederiksen H, Kranich SK, Jørgensen N, Taboureau O, Petersen JH, Andersson AM. (2013). Temporal variability in urinary phthalate metabolite excretion based on spot, morning, and 24-hour urine samples: considerations for epidemiological studies. Environ Sci Technol. (47):958−967.

Saravanabhavan G, Guay M, Langlois E, Giroux S, Murray J, Haines D. 2013. Biomonitoring of phthalate metabolites in the Canadian population through the Canadian Health Measures Survey (2007-2009), Int J Hyg Environ Health. 2013 Feb 15. doi:pii: S1438-4639(12)00145-9.10.1016.

The CHAP used exposure scenario-based analyses (1) as a cross-check of its DI calculations based on the spot urine data; and (2) to understand the possible relative contribution of various phthalate sources to urine concentrations. The CHAP elected not to use these analyses to quantify risk and relied instead entirely on HBM data to calculate margins of exposure (MOEs) and HIs. This approach is reasonable given the generally superior nature of the HBM data for identifying the type and relative contribution of each phthalate to mixtures. However, the exposure scenario-based assessment is the only option available to quantify exposures that have not been measured in urine or that may happen under "foreseeable use and abuse conditions" as specified in the CHAP's charge but that have not yet occurred. Therefore, some consideration of this approach, and recommendations made for improving it, are warranted. The CHAP looked to the future appropriately in identifying several phthalate substitutes and recommending that they be included among chemicals monitored in NHANEs and in future exposure studies. However, new substitutes could emerge over time and CPSC will need a sufficient mechanism for anticipating and preventing any problematic exposures before they occur. This topic warrants some discussion and a recommendation from the CHAP.

The report would benefit from a more complete discussion of trends in phthalate use over time (i.e., types of phthalates/mixtures and their applications) and uncertainties in identifying such uses to (1) better understand and interpret HBM data; and (2) determine whether the CHAP's evaluation is adequately protective of exposures under foreseeable future use patterns. A note at bottom of page 25 indicates that HBM data from NHANES sampling rounds previous to 2005-6 were not used because of study design changes associated with fasting requirements. Another reason to exclude data from earlier rounds might be that they do not reflect current exposures. Does work in Canada (e.g., Saravanabhavan et al. 2013) have any relevance in determining the uncertainty associated with collection of HBM data prior to CPSIA bans? For the scenario-based assessment, the CHAP indicates that its overall goal "was to obtain phthalate related data from the U.S. that were published in the last ten years and use the data to estimate inhalation, ingestion, and dermal exposures to phthalates from contacts with children's toys, and other sources/products" (Page 40, lines 1348-1350). Why ten years? To approximate exposure reflected in HBM data as well as exposure since the HBM data were collected? It would be helpful to explain the temporal aspects of both measures of exposure (i.e., the HBM data and data supporting the scenario-based exposure modeling) to facilitate their comparison by the CHAP and an understanding of their relevance to current and possible future exposure to phthalates.

2.1.1.4 Reviewer 4:

The analysis of biomonitoring data is a step in the right direction in improving exposure assessment. It is especially useful along the line of total exposure through multiple media and routes as well as cumulative exposure to multiple chemicals. The equation for computing daily intake of a phthalate (Section 2.5.3 and Appendix D) is a good approximation, but it depends heavily on population parameter such as the molar ratio between the amount of metabolites excreted in urine and the amount of parent compound taken up. These parameters are subject to uncertainty and variability. More sophisticated (physiologically-based) models are inevitably necessary to better understand the pharmacokinetics of phthalates. To this end, a review or discussion on physiologically-based pharmacokinetic models for phthalates is absent from this Report.

2.1.2 Question 1b: Is the use of spot urine samples appropriate for estimating population exposure?

2.1.2.1 Reviewer 1:

Given the relatively short half-lives of most of the phthalates under investigation, the use of spot urine samples is far from ideal. As the report points out, we know that the variability in urine levels of metabolites can be extremely large on an individual basis which does lessen confidence in the information. Having large exposure datasets, which is not common for many environmental agents, goes some way to offset the variability. This is the nature of the data available and therefore the Panel needed to use their professional judgment on using the best available. The European and NHANES data have attempted to account for potential contamination issues which many of the earlier biomonitoring studies did not. In conclusion, the use of spot urine samples is probably a reasonable approach, given the other exposure data that were available to the Panel. However, are they really getting at fetal exposure subsequent to pregnant female exposure? Some discussion of different metabolic capabilities between dam and fetus and the potential impact on toxicity given current knowledge needs to be added.

2.1.2.2 Reviewer 2:

In my opinion, it is better than nothing, but is more useful as a qualitative indicator that exposure occurred, rather than a quantitative indicator (see continued comments under Question 1.c below). This issue is not evenly discussed in Chapter 2. p. 17, which focuses on reproductive effects from epidemiological studies, provides no caveats about spot urine measurements in the text, whereas caveats are given under neurodevelopmental outcomes (L988). This issue is discussed much better in Chapter 4 (L1861ff) and that information should be discussed more thoroughly in Chapter 2 or the

Chapter 2 material be moved to a later chapter.

2.1.2.3 Reviewer 3:

Generally yes, and I am not aware of better data to support the CHAP's analyses. See the response to Charge Question 1a (see 2.1.1.3).

2.1.2.4 Reviewer 4:

2.1.3 Spot urine samples are certainly useful in looking into short history of exposure (e.g. 24 hours). Utilizing repeated spot urine samples helps us to better understand variation during the day. Urine samples are known for considerable within–subject and between-subject variability even with creatinine adjustment. Use of repeated spot samples from a single subject could better reveal these variabilities and population exposure.

2.1.4 Question 1c: Is the use of spot urine samples likely to underestimate or overestimate the median or upper bound exposure?

2.1.4.1 Reviewer 1:

It seems that this is likely to underestimate the median and upper bounds because of the issues raised in the draft report. One would assume that some attempt was made to control when the samples were collected (e.g. first morning void), but given that exposure could come from multiple sources – with the most likely being food, then it is still difficult to know what true exposure levels were that resulted in the urine measurements obtained because of timing issues.

2.1.4.2 Reviewer 2:

In the case of intermittent exposure, the values could over or under-estimate exposures. They could also provide false negatives if significant exposures occurred in the past, but the chemical was excreted by the time of measurement. The issues with a window of vulnerability during pregnancy complicate interpretation of biomonitoring data. For example, did exposure occur during the window of vulnerability? On the other hand, if the phthalate reached a pharmacokinetic steady-state and the biomarker of the particular phthalate metabolite was representative of the concentration of the parent, then a spot urine sample would be a more accurate indicator of dose. This issue should be discussed more thoroughly in the text. This discussion doesn't arise till p 63, where it is well done.

2.1.4.3 Reviewer 3:

As noted in response to Charge Question 1a (see 2.1.1.3), there are many factors contributing to uncertainty in the application of spot urine concentration data to the questions that the CHAP must address, with some potentially leading to overestimates

of exposure and others leading to underestimates. Without further discussion and quantification of these sources of uncertainty where possible, it is difficult to reach a conclusion.

2.1.4.4 Reviewer 4:

Urine sampling is highly variable, creating uncertainty in data analysis and reduces statistical power. As long as analytical methods are unbiased, the use of spot urine samples itself will not likely to dictate over or under-estimation of exposure.

2.1.5 Question 1d: Does the report adequately characterize the uncertainty of the biomonitoring data and approach?

2.1.5.1 Reviewer 1:

Yes. The report does acknowledge the many shortfalls in the approach, but correctly indicates that they used the best datasets available to them.

2.1.5.2 Reviewer 2:

No. The uncertainty is not discussed until p. 63, where it is well done. Many readers will not get this far and will believe the numbers are far more precise than they are. Biomonitoring data are fundamental to this risk assessment and can/should be utilized. However, the discussion must be balanced, including both strengths and weaknesses. It appears that global (especially German) HBM were used to identify daily intakes. If so, this is of great concern without some kind of evidence for similar exposures (considering different sources and pathways and behaviors) by age group. Section 2.5 describes no weaknesses and overemphasizes the strengths. More specifically:

1) L1031 says HBM determines human exposures. It only determines body burdens at the time of measurement. See comments under c above.
2) The NRC (2006) has a rather extensive discussion of strengths and limitations of HBM, with specific attention to phthalates.
3) L1048 says "HBM data can be used to quantify overall phthalate exposures, to compare exposures of the general population with special subpopulations…and with toxicological animal data." This is a significant overstatement with no balance. The greatest problem is that HBM is used synonymously with "exposure". Exposure is the contact of the person and agents over a specific period of time, which is quite different from body burden.

They can be good (or poor) indicators that exposure occurred. In the case of most phthalates, I think they are good indicators of exposure. I agree that they can be used to *compare* the general and specific subpopulations, although this has significant uncertainty due to different exposure sources and pathways in different age

subpopulations. To compare them with animal toxicology data requires high quality extrapolation modeling to compare HBM to animal biomonitoring data, which apparently wasn't done. Apparently, HBM were modeled to approximate human exposure, which was then compared to the RfD based on exposures used in animal toxicology studies. The better comparison would be to compare human biomonitoring data to animal biomonitoring data. I don't know whether the toxicology studies used for the RfDs did this.

4) Much attention is appropriately given to HBM data. Please indicate the date of the NHANES and SFF biomonitoring data vis-à-vis the bans and interim bans. This is buried in one of the Appendices but should be brought forward briefly. Then discuss the potential impacts of any date differences. It would be particularly important if the HBM data for the banned phthalates remained high after the ban. It would have implications to approaches to reducing risks.
5) The quality of NHANES HBM data is well known. Samples are probabilistic and the measurements are of highest quality. The probability sampling of NHANES is described in Appendix D. The SFF samples were measured in the Centers for Disease Control and Prevention (CDC) labs and therefore were of the highest quality. It does not appear that they were collected from a probabilistic sample (or at least there was no description of statistical elements of the sampling in Appendix D). Please clarify this point, and if not probabilistic, please provide a discussion of the strengths and limitations of using this type of non-probabilistic data for this risk assessment.
6) It appears that HBM data collected outside the U.S. were used in the assessment. For example, L1106 and 1107 include German data (e.g., Koch *et al*., 2003a; Wittassek *et al*., 2011). Tables 2.5, 2.6 and 2.7 have global data. I suggest eliminating the global data from the text and tables, leaving a general statement in the text (with references) that the global findings are "similar" to U.S. data. To use global data quantitatively (even as done on L 1106 for DI calculations) would require far more information about their quality and relevance. For example, whether probability sampling was used, were measurements up to CDC standards, how similar/different are exposure sources and pathways, what is the relationship of European and Asian bans to the time of urine sampling, etc.
7) L1130 refers to data in tables 2.5 and 2.6 estimated by the CHAP. This needs to be referenced. How did CHAP do the estimates? What is "weighted"? I guess this is from Appendix D, but the text should cross-reference the specific section.
8) L 1152ff. Says that "infants might have significantly higher

intakes…compared to their mothers", referring to Figure 2.3 on p 49. This is true, but the infants are similar to or less than toddlers so this should be mentioned.

L1929 says "[HBM] data therefore provides a reliable and robust measure of estimating the overall phthalate exposure and resulting risk." The immediately preceding material is well done and does not support the strength of the words "reliable and robust". It is certainly useful and usable. But I wouldn't characterize these degrees of variably and reliability as robust.

2.1.5.3 Reviewer 3:

For the most part, yes, except as otherwise described in response to Charge Question 1.a (see 2.1.1.3)

2.1.5.4 Reviewer 4:

The Report demonstrates the variability of the biomonitoring data through separate analysis of data for general population, pre-natal and postnatal women, infants using NHANES and SFF data. Knowledge of this variability is helpful. By considering sampling weights in the analysis of NHANES data, the results are statistically generalizable to the US population. The back-estimate of daily intake based on phthalate metabolites is important, but was not really discussed in the Report. However, the Report compared the results between NHANES and SFF studies as well as two model-based estimates. The moderate differences that are within an order of magnitude somewhat validate these estimates. Within this very specific context, the Report looked into uncertainty of the biomonitoring data. Further investigation and better characterization of uncertainty remains desirable.

2.2 Cumulative Risk Assessment

The CHAP calculated hazard indices (HIs) for individuals exposed to multiple phthalates, and then generated distributions of the hazard index. This was done to account for differences in pharmacokinetics and potency among different phthalates. This is also necessary to estimate upper bound risk accurately, that is, to avoid summing 95th percentile exposures from individual phthalates.

2.2.1 Question 2a: Is this approach to cumulative risk assessment appropriate and scientifically defensible?

2.2.1.1 Reviewer 1:

The HI approach seems reasonable and is a tried and trusted method for dealing with mixtures of chemicals (e.g., at Superfund sites). My comment would be more on the determination of the point(s) of departure. Why, when many of the studies have good

dose-response estimates did the Panel choose no observed adverse effect levels (NOAELs) rather than benchmark doses (lower confidence limit) (BMDLs)? The NOAEL is an accident of dose selection, not necessarily a reflection of the uncertainty or steepness (or otherwise) of the response from all the dose levels selected on the studies.

2.2.1.2 Reviewer 2:

In general, yes, for the reasons presented in the document. It is especially valuable to be able to understand medians and upper bounds. However, it needs to be discussed in a more balanced way. More specifically:

1) P17 L844ff (Cumulative Exposure Considerations) This is an important discussion, making attention to detail important.
2) L865 correctly states that the purpose of these studies "was not to investigate the effect …at realistic exposures…Rather, their merit is in demonstrating that mixture effects of these substances can be predicted quite accurately…" I trust the authors in citing the results of the work well. However, I don't believe that mixtures at excessive doses can *accurately* predict mixture interactions in the real world.
3) L868 says "predicted quite accurately…" This is an overstatement. In my opinion, it would be more correct to say something like "can reasonably be predicted." My opinion is based on the lack of realistic exposures used in the underlying studies.
4) L872 has another example of a sweeping statement unsupported by the literature. Indeed, SOME phthalates act in concert with SOME other antiandrogens.
5) L878ff. Not being familiar with the phthalate literature, I read the abstract of the Hotchkiss *et al.* 2004 paper cited. The abstract says that responses were "dose additive rather than synergistic…" this paragraph implies synergism.

L1529 is too sweeping ("human biomonitoring determines internal exposures…") for reasons noted elsewhere.

2.2.1.3 Reviewer 3:

The novel approach to calculating HIs is defensible in that it provides a sound basis for evaluating actual rather than hypothetical phthalate mixtures to which individual members of the U.S. general population are exposed. This approach avoids errors resulting from quantitative assumptions about correlations among phthalate exposures or lack thereof. A limitation of this approach is that there is some uncertainty about the extent to which DIs estimated from single spot urine samples are comparable to the

NOAELs and Reference Doses with respect to the timing of exposure. Ideally, repeated spot urine samples would be collected from each individual to obtain a temporal average over the exposure duration of concern given the toxicity and degree of exposure to phthalate mixtures. It is possible that variability of the HI distribution is larger than it would be if each HI had been calculated from biomonitoring data that better represented average daily exposure over the exposure duration of interest for phthalate mixtures. This topic warrants some discussion along with other sources of uncertainty in the HBM data and their application as noted in response to Charge Question 1.a.

By convention, HIs are reported with one significant figure given the uncertainty associated with the exposure and toxicity information underlying these values. However, some of the HIs in Table 2.16 are reported with as many as three significant figures. Given the uncertainty underlying the CHAP's exposure and toxicity assessments, it is not apparent that this level of precision has really been attained. Also, three HIs do not match the values presented in Table D-9 and appear to be in error: (i) NHANES Case 2 is listed as 7.4S, which is listed as 7.4 in Table D-9; (ii) SFF infants Case 1 is listed as 34.7, which is listed as 3.71 in Table D-9; and (iii) SFF infants Case 2 is listed as 2.39, which is slightly different from the 2.32 listed in Table D-9.

2.2.1.4 Reviewer 4:

The use of HI is interesting and useful. Its validity hinges on the dose-additivity assumption, which can be a reasonable one often when the multiple phthalates are of very low amounts. But more fundamentally, this assumption may imply that all phthalates in the mixture share something in common, whether it is concerning pharmacokinetics, mechanism of action, or chemical property, which is more difficult to verify. Therefore such an assumption needs to be checked and verified on a case-by-case basis for a different mixture of phthalates. In the absence of interaction among element phthalates, HI does provide a practical way to measure the cumulative exposure and associated risk.

In the Chapter 5, however, CHAP's recommendations do not appear to be closely related to HI.

2.2.2 Question 2b: Are there alternative approaches you would recommend that the CHAP consider?

2.2.2.1 Reviewer 1:

See response to Question 1a (2.1.1.1) - estimate BMDLs for the POD and then use the HI approach.

2.2.2.2 Reviewer 2:

No.

2.2.2.3 Reviewer 3:

No, not with the data currently available to the CHAP. Given the complex pattern of consumer product use and variability in phthalate use over time, biomonitoring measures for a given individual remain the best indicator of the mix of phthalates to which they are exposed.

2.2.2.4 Reviewer 4:

A similar approach in general methodology is the so-called response-additivity assumption (EPA 2000). When there are interactions (either synergism or antagonism), the situation is complex, and research is on-going.

2.3 Critical Effect and Reference Doses

2.3.1 Question 3a. The risk assessment focuses on male developmental effects. Is the choice of male developmental effects as the critical effect supported by the available animal and human data?

2.3.1.1 Reviewer 1:

Probably. The Panel does provide an explanation for the choice of the male reproductive developmental end point and clearly this also is the only one where empirical evidence exists on mixtures *in vivo* showing dose additivity – one of the charges to the Panel. The Panel also relied heavily on the NRC report, but I think other toxicities could have been given some greater discussion. How and why was this selected as the primary toxicity? Was it based on dose effect levels in toxicity studies? published RfDs? sensitivity of the end point? The document needs a larger rationale for selection or perhaps a reference to others who have done this as well for concordance. So for example, the Panel gave a cogent argument for why they did not consider the cancer end points noted for many phthalates. Much of this was based on the IARC review of DEHP (IARC, 2000) with the notion that the liver tumors noted were likely driven by a PPARα-mediated mechanism that may not have relevance to humans. There are some data available that DEHP still produces tumors in a PPARα – knock out mouse (Ito *et al.* (2007) suggestive that more than one mechanism of hepatocarcinogenesis may be operating for DEHP, but also ignoring the data that DEHP also produces testicular Leydig cell tumors (Voss *et al.*, 2005) and pancreatic acinar cell tumors (David *et al.*, 2000) in the rat, that have not been causally related to PPARα activation. The former is also interesting in that Leydig cell tumors have also been noted after *in utero* only exposure to phthalates (Barlow and Foster, 2003;

Mylchreest *et al.*, 1999) and may be part of the testicular dysgenesis syndrome. While some publications have noted that human Leydig cell tumors are extremely rare (Cook *et al.*, 1999) others have indicated that this may be due to a difference in pathology diagnosis (Foster, 2007; Holm *et al.*, 2003) and that human Leydig cell micronodules would likely be diagnosed as Leydig cell adenomas, if examined by a veterinary, rather than a medical, pathologist and that such micronodules are common in infertile men.

So while there are good, pragmatic reasons for the selection of the end point chosen, I think the cancer end points deserve a greater consideration and emphasis by the Panel than what is currently available in the document.

2.3.1.2 Reviewer 2:

This is appropriate. However, there should be an expanded discussion of the weight of evidence (WOE) from the epidemiology studies (e.g., p 19). The discussion of the study results is good. However, for those not familiar with these particular studies or the utility of epidemiology studies in general, it would be useful to discuss the strengths and limitations of the three study cohorts used (e.g., statistical power, biomonitoring protocols) and well as the use of epidemiological data to show association vs. causality. Table 2.2 is particularly good. This is not done until p 64ff.

2.3.1.3 Reviewer 3:

Yes. This question was examined in detail by the 2008 NRC committee that recommended performance of cumulative risk assessment for phthalates. Some members of that committee also serve on the CHAP, and the CHAP has built upon the NRC committee review by evaluating (1) peer-reviewed scientific studies published since the NRC committee completed its review; and (2) unpublished, ongoing research regarding phthalates and the applicability to humans of the male developmental endpoints observed in animal studies. This unpublished work presented to the CHAP raised questions about whether male developmental effects represent the critical effect in humans, but I concur with the CHAP that the information provided thus far, at least as it has been described in the CHAP's draft report, is not sufficient to refute the applicability of the rat model to humans.

The CHAP could more clearly defend its choice of critical effect by tabulating NOAELs across endpoints for each phthalate to show that, where quantitative comparisons are possible, NOAELs associated with male developmental effects are lower than NOAELs associated with other endpoints. This comparison might be particularly helpful for neurodevelopmental effects. Comparisons also might be of interest to adult men and women who are not of child-bearing age who are wondering about the extent to which they should be concerned with their own exposure to

phthalates.

2.3.1.4 Reviewer 4:

There is solid evidence to support the existing modes (mechanisms) (e.g. anti-androgen) for male reproductive and developmental effects based on animal studies, but human data are lacking. While it is adequate to focus on reproductive and developmental effects only in risk assessment, regulatory decisions take all likely health effects into consideration, particularly the most sensitive ones and most susceptible populations.

2.3.2 Question 3b: Is it appropriate to regard male developmental effects in rodents as the critical endpoint for the cumulative risk assessment of phthalates in humans?

2.3.2.1 Reviewer 1:

Probably, but see comments in 2.3.1.1 above in the dealing with a more comprehensive assessment of other toxicities.

2.3.2.2 Reviewer 2:

It is appropriate to use such effects for extrapolation and to use the rat because rats are the most sensitive species. However, I have concerns about the quality of the quantitative extrapolation to humans (see responses below under question 3c, 2.3.3.2).

2.3.2.3 Reviewer 3:

Yes. The CHAP provides a clear and compelling argument for choosing this critical endpoint. The CHAP summarizes the evidence of such effects in epidemiological studies, notably reduced anogenital distance associated with gestational exposure to DEP, DBP, and DEHP, and how the evidence compares with findings from studies of male rats (page 19, lines 961-975).

2.3.2.4 Reviewer 4:

Within the context of risk assessment of overall health effects, it is important that we evaluate the reproductive and developmental effects that are supported by plausible mode of actions. At the same time, risk assessment also must reflect all possible health effects, especially the more sensitive ones. Towards that end, it is important to consider all health effects, especially in view of the impact of any policy change on potential effects of phthalates.

2.3.3 Question 3c: Is there sufficient understanding of the phthalates' mode of action to extrapolate from male rat developmental effects to humans?

2.3.3.1 Reviewer 1:

There is significant information on some of the critical effects for initiation of the

developmental effects on the male rat reproductive system. For example, we know that there are at least three distinct MOAs operating to produce the syndrome of responses associated with *in utero* exposure to specific phthalates (1) lowered fetal testicular testosterone production (related to malformations of the male reproductive tract – such as the epididymis, vas deferens, prostate, etc.);(2) lowered production of the Leydig cell product *insl3* (related to the development of the gubernaculum and development of cryptorchidism) and (3) effects on fetal Sertoli and germ cell function (e.g., by changes in *ckit* expression and stem cell factor production) leading to the generation of multinucleated gonocytes in the fetal testis. Information on (1) and (2) is fairly comprehensive and for (3) is minimal. We do know that significant decreases in fetal androgen production in the human fetal testis will lead to malformations of the reproductive tract. There is also evidence that issues with *insl3* receptor expression in human fetuses are associated with cryptorchidism (covered in the NAS review, NRC, 2008). So similar mechanisms do operate in various human congenital syndromes, but direct evidence for interactions with phthalates leading to reproductive tract malformations is not present, although some specific associations do exist, with for example, Anogenital Distance (AGD), neonatal testosterone levels.

2.3.3.2 Reviewer 2:

Yes, because it is essential to assess risk "now". However, the understanding is quite poor. Perhaps the greatest problem is the large interspecies differences in responsiveness. For example, is a human more like a rat or more like a mouse? Prudence and standard risk assessment practices clearly indicate that using the rat is appropriate, but it must be understood that future information could question the use of rats. The discussion of this issue on p. 7 should be expanded. For example, in L483ff, the doses used in the studies are not given. The Gray *et al*. (1982 used 2000 mg/kg/day for 7 or 9 days. The Gaido *et al*. (2007) used 250 or 500 mg/kg/day. The paragraph on L504ff (nonhuman primates) describes the dosing well. All these doses are extremely high relative to human exposure. This raises a serious question of whether the data can be reliably extrapolated at all to a human exposure scenario. Specifically, did these doses overwhelm normal metabolic clearance mechanisms, in which case the dose-response would not be linear? One crucial element missing from the document is a lack of discussion of pharmacokinetics and the potential to overwhelm clearance mechanisms. The discussion is better in Chapter 4.2, but still would benefit from some expansion.

More care is needed in summary statements. For example L639 says "These findings are most prominent in rats although inconclusive studies in humans suggest that similar effects may be seen in humans." I agree with this sentence up to the word similar. To extrapolate *all* of the rat phthalate syndrome to humans is too strong. This could be

fixed by adding the word "many" before the word "similar". "May be seen" is also way too strong given the weaknesses in understanding interspecies pharmacokinetics, MOA, WOE, and the epidemiology studies. It implies that the effects may actually be observed. The language should be more caveated, although the possibility needs to be stated.

2.3.3.3 Reviewer 3:

First, the question implies that phthalates exhibit a single mode of action in the disruption of male development. In reality, they might have more than one mode of action, but the multiple modes could contribute to common adverse health outcomes. There is a reasonably good understanding of modes of action in rats, but less so in humans. Nevertheless, as noted in response to Charge Questions 3.b, the evidence that male rat developmental effects might be relevant to humans is compelling.

2.3.3.4 Reviewer 4:

MOA is outside of my field. My impression through reading the literature is that there seems a consensus on the mode of action of phthalates especially concerning male reproductive and development effects.

2.3.4 Question 3d: In case 1, is the selection of published reference doses for individual phthalates (Kortenkamp and Faust, 2010) appropriate for this task? (See section 2.7.2.2 of the CHAP draft).

2.3.4.1 Reviewer 1:

As a case study, this is a reasonable approach, but again I would have wished to see a better estimate of the dose response via BMDL rather than NOAELs for this activity.

2.3.4.2 Reviewer 2:

Firstly, what is the definition of "reference dose?" The dilemma is that the Kortenkamp and Faust (2010) paper is appropriate for this task simply because of the very broad sweeping risk assessment; i.e., more precision in the RfD would not change the conclusions. The Kortenkamp and Faust paper is flawed because it provides neither a description of how the UFs were decided upon nor any description of how/whether the animal-to-human extrapolation was done. To be of quality, the animal-to-human UF should include consideration of pharmacokinetics (PK) and pharmacodynamics (PD). The PK should be calculated. The PD might need to rest on a 3 UF. Furthermore, L1501 says that the RfDs of EPA and Acceptable Daily Intakes (ADIs) of CPSC were used along with other RfDs. Perhaps it is there, but I did not see any EPA or CPSC values. The EPA IRIS values that I checked were quite out-of-date. A summary description in Table D-8 says Kortenkamp and Faust applied an UF of 500 to DINP because only a lowest observed adverse effect level (LOAEL) was available. The

typical LOAEL to NOAEL UF is 10.Use of a combined UF of 500 is quite unusual because UF are 1, 3 or 10, not 5. Nevertheless, when the RfD process is described, it should be indicated that more precision through PK is desirable, but not practical and meaningful given the overall database.

2.3.4.3 Reviewer 3:

In response to questions 3.d and 3.e, generally, yes. Kortenkamp and Faust (2010) are among the few scientists investigating cumulative exposure to phthalates and its possible effects. The CHAP did not stop with this work but evaluated relevant data de novo to develop its own RfDs. Some of the RfDs used to quantify HIs are based on benchmark doses (BMDs). The CHAP identified NOAELs but does not discuss whether any dose-response data were amenable to BMD modeling and, if so, whether such modeling might lead to different conclusions for any individual phthalate. The CHAP was charged with evaluating many phthalates for many toxicity endpoints, but some discussion of BMD modeling options is warranted.

The CHAP is also largely silent about how it arrived at the uncertainty factors applied to NOAELs to estimate RfDs. For the five phthalates with HI estimates, the CHAP assigned total uncertainty factors of 100, with a factor of 10 to account for inter-species extrapolation and another factor of 10 to account for inter-individual variation (page 53, lines 1591-1592). The CPSC charge requires the CHAP to

> "consider the level at which there is a reasonable certainty of no harm to children, pregnant women, or other susceptible individuals and their offspring, considering the best available science, and using sufficient safety factors to account for uncertainties regarding exposure and susceptibility of children, pregnant women, and other potentially susceptible individuals."

The CHAP's work is extensive, detailed, and meticulous in so many regards, so it is puzzling that such an important component of the analysis relies on standard default values without discussion of relevant literature. Quantitative data-derived uncertainty factors specific to the chemical of interest have been discussed in the scientific literature for some time by scientists (e.g., Dale Hattis, Mike Dourson, Sandra Baird, and numerous others), and the U.S. Environmental Protection Agency is developing guidance for developing data-derived uncertainty factors specifically for interspecies and intraspecies extrapolation (http://www.epa.gov/raf/DDEF/index.htm).

The CHAP quantified a range of RfDs for individual phthalates. It is interesting to compare these RfDs to those currently being used around the world as in indicator of

how the CHAP's updated review of the toxicological literature might influence risk assessment of these phthalates. The following table provides this comparison. Current RfDs established by the U.S. Environmental Protection Agency and other international entities are either not available or within the range of those used/developed by the CHAP except for DEHP. An assessment conducted in the Netherlands resulted in an RfD of 0.004 mg/kg-day, which is about seven times lower than the low end of the CHAP's range of RfDs; however, this lower RfD is based on a relatively old assessment completed in 2000.

Phthalate	**Range of RfDs from Cases 1, 2, and 3 (mg/kg-d)**	**Range of RfDs from other sources (mg/kg-d)**
DIBP (diisobutyl phthalate; CAS # 84-69-5)	0.05 to 1.2	not available
DBP (di-n-butyl phthalate; CAS # 84-74-2)	0.05 to 0.5	ATSDR: not available Health Canada: 0.063 (1992; mouse fetotoxic/teratogenic) RIVM: 0.052 (2000; rat embryo toxicity) USEPA IRIS: 0.1 (1987; rat increased mortality) USEPA PPRTV: not available
BBP (butylbenzyl phthalate; CAS # 85-68-7)	0.05 to 0.5	ATSDR: not available Health Canada: 1.3 (1998; rat pancreas) RIVM: 0.5 (2000; rat/mouse kidney, haematopoietic system, testes) USEPA IRIS: 0.2 (1989; rat liver) USEPA PPRTV: not available
DEHP (di [2-ethyl-hexyl] phthalate; CAS #117-81-7)	0.03 to 0.05	ATSDR: 0.06 Minimal Risk Level; MRL (2002; rat testes) Health Canada: 0.044 (1992; mouse/maternal and fetal) RIVM: 0.004 (2000; mouse testes) USEPA IRIS: 0.02 (1987; guinea pig liver) USEPA PPRTV: not available
DINP (diisononyl phthalate; CAS # 28553-12-0 and 68515-48-0)	0.115 to 1.5	Not available

ATSDR – Agency for Toxic Substances Disease Registry
RIVM – Netherlands National Institute for Public Health and the Environment
USEPA IRIS – U.S. Environmental Protection Agency Integrated Risk Information System Database
USEPA PPRTV – U.S. Environmental Protection Agency Provisional Peer-Reviewed Toxicity Values

Source: International Toxicity Estimates for Risk (*ITER*) database at the National Library of Medicine's Toxnet (http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?iter)

The CHAP appropriately acknowledges inhalation and dermal pathways, and exposure from these pathways is evaluated in the scenario-based assessment. However, the CHAP did not quantify RfDs and Reference Concentrations (RfCs) specific to these pathways. It did not need to do so because the scenario-based assessment did not include quantification of risk from these pathways. Also, the relevant animal toxicity

studies involve oral exposures to phthalates. However, it would be useful to note that toxicity values specific to the evaluation of the dermal and inhalation pathways are not available.

2.3.4.4 Reviewer 4:

Yes. The RfDs are defined using conservative uncertainty factors.

2.3.5 Question 3e: In cases 2 and 3, is the derivation of the individual phthalate reference doses appropriate, including selections of studies, endpoints, and uncertainty factors? (See section 2.7.2.2 of the CHAP draft).

2.3.5.1 Reviewer 1:

All seem very reasonable in illustrating specific case studies – similar criticism to that provided in Question 3.d (see 3.3.4.1) would also apply.

2.3.5.2 Reviewer 2:

These RfDs have the same problems discussed in Question 3.d above (see 3.3.4.2). It is not possible to answer this question because the explanations are inadequate and sometimes confusing. In many cases, information is just stated with no reference or cross-reference elsewhere in the document. In other case, a cross reference to another location is given, but this other location has an inadequate description. Some specifics follow:

- Often an UF of 100 was applied. The 100 was derived from 10 for interspecies and 10 for intraspecies. However, apparently no PK extrapolation was performed. Although such extrapolation is scientifically desirable, it would not change the end results because of the large uncertainties involved.
- Why does Table 2.15 have columns for POD, rather than calling them NOAELs which they most often are called in the text?
- L 1587ff. Case 2 hinges on the NOAEL for DEHP. Table 2.15 selects a POD (NOAEL) of 5, without explaining why the 5 since Case 1 has a different one (3). Possibly, it was the consensus NOAEL of Case 3. The point is that an explanation is needed.
- More specifically, the Cases in Section 2.7.2.2 are described in a total of 33 lines plus Table 2.15. I found additional descriptions in Appendix D (L 218 to 222). Here, the descriptions were even briefer. There are adequate descriptions of the study NOAELs. However, in many cases the UFs are not adequately described (see response under Question 3d; 3.3.4.2 above). Appendix A has a very good discussion of the various papers and the rational for the consensus NOAEL.
- L1594 begins a 2-sentence description of the methods used for Case 3. It says that the RfD anti-androgenicity values are in Table 2.1. Table 2.1 (p15) has no

RfDs. Then it refers the reader to Table 2.15, which does have the RfDs. However, there are neither references nor an explanation of why the RfDs are so different. For example, why are the PODs so different? Were different studies used? If so, why? After much looking, I found a description of the papers used for the Case 3 POD in Appendix A. Appendix D (L564ff) refers back to Section 2.3 which refers to Appendix A. Appendix A is good and is clear. It appears that Appendix A "consensus NOAELs" were derived by CHAP, which means they are for Case 3. A comparison of Table 2.15 POD to the consensus NOAELs in Appendix A reveals a discrepancy for DINP (table 2.15 says 50, whereas Appendix A, L723 says 300).

2.3.5.3 Reviewer 3:

See Reviewer 3 response to Question 3.d (2.3.4.3).

2.3.5.4 Reviewer 4:

As long as these NOAELs are derived using a system that is clearly-described, they serve for at least one purpose: to demonstrate the uncertainty and sensitivity of HI to the choice of RfDs (PODs). Within this context, the use of various choices should be explored. However, there seem to be inconsistency in these NOAELs. For example, the NOAELs reported in Table 2.1 cannot be located in the Appendix (Table A-8 and A-10). Further, for most the phthalate substitutes and a few not-banned or interim ban phthalates the NOAEL is based on non-anti-androgenic effects (e.g. systemic effects), which is fine but the Draft states for instance in case 3 the NOAELs are associated with reproductive and developmental endpoints (line 564-566).

2.3.6 Question 3f: Are there other endpoints that should be considered for risk assessment (either for individual phthalate risk assessments or cumulate risk assessment)?

2.3.6.1 Reviewer 1:

While I think the Panel has undertaken a very reasonable approach in selecting the critical endpoint, I believe some of the cancer data do warrant some further investigation – probably on a specific phthalate basis since it is unlikely that studies have been conducted for all the phthalates under review and potentially in strains or species that are not helpful for the targets of testis , liver or pancreas.

2.3.6.2 Reviewer 2:

L1562 Says that "despite human studies…MEP…reproductive…, these phthalates were not considered…" Please explain why.

2.3.6.3 Reviewer 3:

For the purpose of reaching decisions about use of phthalates in consumer products to which children might be directly or indirectly exposed, the selected endpoint should be sufficient to protect fetuses and young children. However, evaluation of other endpoints could be helpful in providing risk-based advice to other members of the population (i.e., women of non-childbearing age and men, the elderly, etc.). And, of course, the CPSC should continue to follow the scientific literature for new information that would suggest the need for new bans or the elimination of existing bans.

2.3.6.4 Reviewer 4:

The Draft tends to base its derivation of NOAEL on the most sensitive reproductive and developmental effects (with a lower NOAEL value). Within the context of risk assessment, it helps to differentiate the situation where the evidence of reproductive and development effects may be absent, but that of other potential health effects (e.g. systemic effects) is present. So the final risk estimate may be driven by a non-anti-androgenic effect.

2.4 Sensitive Populations

2.4.1 Question 4a: Is the selection of sensitive populations appropriate?

2.4.1.1 Reviewer 1:

See Reviewer 1 response to Question 4b (2.4.2.1 below).

2.4.1.2 Reviewer 2:

Yes.

2.4.1.3 Reviewer 3:

Given the critical endpoint, yes.

2.4.1.4 Reviewer 4:

Pregnant women, infants, and children form the most sensitive populations in view of the underlying MOA.

2.4.2 Question 4b: Are "women of reproductive age" the most sensitive population?

2.4.2.1 Reviewer 1:

NO. The most sensitive population would be the male fetus at the critical stage of sexual differentiation. "Women of reproductive age" is a surrogate since the period of development associated with the development and differentiation of the male reproductive tract is relatively early in gestation and may start prior to a confirmation of

a pregnancy. Clearly exposure of the fetus is via the mother, but I do not see a discussion of factors that may indicate that the fetal exposure may be different from the mother (e.g., in exposure to active monoesters rather than conjugates). At the critical window for male sexual differentiation the fetal liver and placenta may not handle phthalate conjugates in the same fashion as seen in the dam and in her circulation, or at later times in fetal development.

2.4.2.2 Reviewer 2:

They clearly are *not*. I agree that the fetus is the most sensitive and that women of reproductive age are appropriately used as a surrogate. There is no real alternative to this terminology, but the surrogacy should be made much clearer in the document. The same comment would be true of "pregnant women." In many cases, the document talks about reproductive and developmental effects in pregnant women, without indicating their surrogacy.

2.4.2.3 Reviewer 3:

Technically, fetuses are the most sensitive population, not the women of reproductive age.

2.4.2.4 Reviewer 4:

Given the short life of phthalates and rodent models that show critical gestation window of exposure, we need additional evidence to think "women of reproductive age" as the most sensitive population.

2.4.3 Question 4c: Does the risk assessment methodology adequately address the potential risks to children?

2.4.3.1 Reviewer 1:

Yes. Moreover there is some delineation of children at various ages. The Panel can work with the rich dataset afforded them (much more so than other environmental contaminants with a short half-life) and have taken a reasonable approach to consolidate the information.

2.4.3.2 Reviewer 2:

Yes. The endpoints used to calculate risk are based on children (fetus-child). HBM data for children are used in the assessment. Appropriate "safety" principles are applied.

2.4.3.3 Reviewer 3:

The CHAP makes a compelling case for fetuses being the most potentially affected life stage for the effects of phthalates on male development, with sensitivity decreasing

with age. Therefore, evaluation of HBM data for women of child-bearing age and infants combined with relevant toxicity data should be protective.

2.4.3.4 Reviewer 4:

There are not sufficient data on humans or children. CHAP argued that rodents are more sensitive than humans and certain reproductive effects are very sensitive to exposure to phthalates. Therefore rodent models combined with sensitive endpoints can adequately protect children. Given the lower exposure level potentially from toys and children's care products, the current approach likely provides adequate protection to children. Note that children have the highest exposure to certain phthalates. Thus, continued monitoring is necessary.

2.4.4 Question 4d: In case 3, the CHAP derived RfDs specific for anti-androgenic effects in male offspring exposed perinatally. These RfDs are not necessarily the most sensitive endpoints for a given phthalate. Is it appropriate to apply reference doses based on prenatal exposure to infants or other populations?

2.4.4.1 Reviewer 1:

Why were only anti-androgenic effects considered? The phthalate syndrome encompasses more than just decreased fetal testosterone levels. There is no mention of *insl3* effects, or effects on fetal testis morphology – multinucleated gonocytes. Can the Panel provide a clearer rationale for this?

2.4.4.2 Reviewer 2:

Not being highly familiar with the literature, I will answer this from a broader principle perspective. If an RfD is protective of the most sensitive endpoint in the most sensitive population, it will be even more protective of less sensitive populations and less sensitive endpoints. For example, if an RfD protects against a statistically significant shift in liver metabolism, that RfD would also protect against mortality. A complexity arises when endpoints differ across susceptible subpopulations. For example, suppose the RfD for an immunotoxic endpoint of an adolescent were lower than for a change in fetal male reproductive tract development. This adolescent may not be protected by action on the fetal RfD. This also illustrates the dilemma of compartmentalization of federal regulations. One may have excellent risk management of one pathway for one population group, but that chemical may present a problem via another pathway for another group. The charge seeks to address total exposure to one chemical and cumulative exposures. In terms of risk management, RfDs are simply tools to be combined with exposure in risk assessment. They would need to be applied differently to different age groups because risk management addresses exposure. So, for example, banning a chemical from a chew toy would have no impact on fetal exposure.

2.4.4.3 Reviewer 3:

In the case of phthalates, it appears to be protective to do so, but perhaps overly protective. However, there do not appear to be sufficient data to develop alternate RfDs. The CHAP states that

> "The RfD values in these cases were derived from *in vivo* evidence of reproductive or developmental effects in pregnant animals. Less is known about the PODs for infants. However, there is evidence that the most sensitive time of exposure is *in utero*, so RfDs associated with reproductive or developmental effects in pregnant women should be protective for infants (Appendix D-9)."

As noted earlier in response to Charge Question 3.f, there is value in being clear about the subpopulation to which the CHAP's range of RfDs (and MOEs) may be most relevant.

2.4.4.4 Reviewer 4:

See Reviewer 4 comments in 2.3.6.4.

2.5 Scope

2.5.1 Questions 5: Did the CHAP adequately address their charge, as outlined above?

2.5.1.1 Reviewer 1:

Yes. I thought that the Panel attempted to keep the report relatively short and readable, but still provide sufficient information on how they operated, and selected appropriate information (and did not select others) and the critical components that led them to their conclusions. They clearly delineated exactly how they were approaching the issues, collecting and marshaling information to address the specific charges made to them as outlined.

2.5.1.2 Reviewer 2:

Yes. One can find answers to all the charge questions within the report. The only limitations are that (1) the communication value of the report should be significantly improved to make it easier to find these answers and (2) the risk management complications of aggregate exposure need to be described more in the main text. See specific comments elsewhere.

2.5.1.3 Reviewer 3:

Except as otherwise indicated in these comments, yes. The CHAP reports that:

> "In an effort to complete its assignment within a reasonable time frame, the CHAP drew some boundaries around the task regarding the number of chemicals to be reviewed, identification of the most sensitive sub-populations, and the endpoint of toxicity of greatest concern" (page 3, lines 319-321)

This statement invites the question - with more time, what other analyses would the CHAP have done?

The CHAP streamlined its work by limiting its review of toxicity information to a subset of all possible phthalates:

> "After careful consideration by the committee, this review is limited to the 3 permanently banned phthalates (DBP, BBP, and DEHP), the 3 phthalates currently on an interim ban (DNOP, DINP, and DIDP), and 8 other phthalates (DMP, DEP, DPENP/DPP, DIBP, DCHP, DHEXP, DIOP, and DPHP)." (Appendix A, page A10, lines 333-336)

Consideration of what exactly? Evidence of male developmental toxicity? Toxicity literature on any form of toxicity? Evidence of current or possible future use in products regulated by CPSC?

In the end, the CHAP quantifies HIs for only five phthalates because of limitations in the HBM data. The report might benefit from a simple flow chart showing the logic behind the CHAP's progression from the very broad scope defined by the CPSC (e.g., all possible effects of all phthalates) to its streamlined scope. The streamlined scope might be a source of concern for some, so the CHAP ideally would be clear about its criteria for defining the scope, along with quantitative supporting analysis where possible, that demonstrates clearly and succinctly why the reduced scope satisfies the goals of the CPSC charge.

2.5.1.4 Reviewer 4:

Charge 1: examine all of the potential health effects (including endocrine disrupting effects) of the full range of phthalates:

The report provides a comprehensive review of all health effects. However, the review process would be more transparent and the results more credible if the process is clearly described and guided by a set of criteria. See the comments regarding systematic

review.

Charge 2: consider the potential health effects of each of these phthalates both in isolation and in combination with other phthalates;

In this report CHAP reviewed the potential health effects, but mostly in isolation of individual phthalates. The lack of studies that were designed to investigate the effects of phthalates in combination hinders the ability of CHAP to investigate the combined effects. The HI approach, under dose-additivity assumption is a useful tool for quantifying the potency of cumulative exposure of phthalates. Whereas there is a very detailed discussion on HI in Appendix D, the discussion is extremely muted in how HI is utilized to derive the CHAP recommendation in chapter 5 of the Report. The application of HI, if any, should be made more explicit.

Charge 3: examine the likely levels of children's, pregnant women's, and others' exposure to phthalates, based on a reasonable estimation of normal and foreseeable use and abuse of such products;

One of the strengths of CHAP report is the analyses of biomonitoring data. The biomonitoring data shows likely exposure levels of phthalates from all sources in different population groups that somewhat represent the US population. The analyses are based on cross-sectional survey. It would be important to continue this analysis to monitor exposure trend over time. This is particularly important regarding phthalates not banned and phthalates substitutes.

Charge 4: consider the cumulative effect of total exposure to phthalates, both from children's products and from other sources, such as personal care products;

The Report presents reasonable estimates of total exposure to phthalates from multiple sources, including that of children's products. Data on children's products remain limited. There appears no animal or human studies that investigate the health effects of total exposure of phthalates.

Charge 5: review all relevant data, including the most recent, best-available, peer-reviewed, scientific studies of these phthalates and phthalate alternatives that employ objective data collection practices or employ other objective methods;

The Report does appear to have a comprehensive review of the existing literature; and presents a summary of the review in Appendices A, B, and C. It is not possible to evaluate the review process as to the degree it has uncovered "the most recent, best-

available, peer-reviewed, scientific studies of these phthalates and phthalate alternatives that employ objective data collection practices or employ other objective methods". To do this, we must know how the systematic the process was and what type of criteria it used. See comments on the approaches taken by the Report.

Charge 6: consider the health effects of phthalates, not only from ingestion, but also as a result of dermal, hand-to-mouth, or other exposure;

This Report cannot answer the question of health effects due to total exposure of phthalates from multiple sources; but it did consider total exposure based on available data. It identifies that the majority of phthalate exposure is from foods; it also identified exposure for children, for example, through hand-to-mouth, bathing product, etc. No data currently exist so that we may infer on risk attributable to each source.

Charge 7: consider the level at which there is a reasonable certainty of no harm to children, pregnant women, or other susceptible individuals and their offspring, considering the best available science, and using sufficient safety factors to account for uncertainties regarding exposure and susceptibility of children, pregnant women, and other potentially susceptible individuals;

CHAP generally followed the paradigm of risk assessment in deriving RfDs. In the process CHAP in fact often took a more conservative approach and chose a consensus NOAEL that is at the lower end of many when available. But inconsistency does occur. For example, DEHT is not viewed as anti-androgenic. As a result, CHAP proposed a NOAEL the highest dose (747 mg/kg-d) that was administered to rats in GD 0-20 (Faber et al. 2007b). However, based on reduction in maternal weight in rats (or mice) the NOAEL for maternal toxicity would be 458 mg/kg/d (or 197 mg/kg-d) for rats (mice). If reproductive effects (i.e., anti-androgenic effects) trumpet other effects, CHAP must make this explicitly and justify this criterion. Risk assessment chooses the most sensitive endpoint to assess overall risk in the worse-case scenario. Even if one NOAEL is chosen to reflect reproductive endpoints only, it must be viewed as a part of overall strategy and in conjunction with protecting other health effects.

Charge 8: consider possible similar health effects of phthalate alternatives used in children's toys and child care articles;

CHAP had made good efforts towards this end. It appears, according to CHAP literature review, there is very little health effects data on phthalates alternatives.

2.6 Overall Conclusions and Recommendations

2.6.1 Question 6: Please comment on the overall conclusions and recommendations

2.6.1.1 Reviewer 1:

I thought the Panel did an excellent job of stating the recommendations and conclusions for the individual phthalates and replacements. These were straightforward, brief, provided clear reasoning for the recommendations and were conservative in the protection of Public Health – I thought that this was very logical and unequivocal in the recommendations to CPSC. It did not require the reader to go searching back through information and were therefore the conclusions were "self-contained". I assume that the recommendations were unanimous or that a consensus was derived in the Panel. I would concur with the overall recommendations based on the information available to the Panel.

2.6.1.2 Reviewer 2:

1) The overall goal of the charge was to identify ways for CPSC to protect children from phthalates and substitutes and to understand this in the context of total risk from total exposure. This is brought out within the main document, but needs expansion to provide clear support for the recommendations. In my view, the most important succinct explanation of this issue is in Appendix E1, L843ff. It clearly says that CPSC can only protect children from a fraction of exposure. For example, the discussion on Section 4.2.1 (L1785) does not explicitly describe the large contribution of diet and other pathways not under CPSC jurisdiction. Although the document provides an excellent analysis of aggregate exposure, it does not use this information in the recommendations. Figure E-1-2 (P E1-38) is quite good in making this point. The concept of thinking about the CPSC-controllable portion of exposure should be thought out and explained well. For example, the document correctly states that DEHP dominates the HI, although it has been banned in CPSC controllable sources. Thus, since DEHP is already a concern because of total exposure, the exposure should not be allowed to increase further and therefore the ban be continued. DIBP (p. 97) provides another example. Based on total exposure data, the MOE ranges from 3600 to 89000 at the 95th percentile and would not be a current concern from that perspective. If this chemical were added to children's toys, the exposure would increase, perhaps into an MOE region of more concern, especially considering cumulative exposure. One can discern these elements from the entirety of the document, but this section on recommendations will be a focus of most readers. Thus, it is important to answer the question: if the MOE is so large now, why worry? Clearly explaining this whole issue may also add impetus to decisions of other regulatory agencies which can do more than CPSC to prevent unacceptable risk from certain phthalates.

2) The risk communication of the conclusions needs significant improvement. Of greatest concern is the footnote on p. 53 that says "When the HI>1.0, there may be a concern for adverse health effects in the exposure population." However, the main document does not adequately describe the RfD as being a significantly "protective/safe" value. As exposure increases above the RfD, concern increases, but it is not a bright line, as implied by this footnote. Also, the exposure values have considerable uncertainty, especially if the SFF data were not probabilistic. A risk manager should have a fuller explanation than can be found in a footnote. Then on L1651, this statement gets stronger, saying that "…HIs that exceed 1.0 are generally considered associated with unacceptable risk…" Unacceptable is a risk management term. The HI is a scientifically based risk assessment term. The difference needs to be clearly drawn. The word "acceptable" and "unacceptable" are quite strong. I would agree that concern increases above a HI of 1, but I would not agree that risk is unacceptable, at for example, 1.1 given the conservative nature of the assessment process. The text on L1689 is far more balanced.
3) I did not find a discussion about linear vs. nonlinear health assessments. This document relied upon NOAELs (or sometimes LOAELs) divided by uncertainty factors. However, many of the studies used exceedingly high dosages, raising the possibility that clearance mechanisms were overrun, resulting in effects that would not occur at lower doses. Even if the data are inadequate to consider non-linear mechanisms, this should be discussed.
4) L1677ff Pages 58 and 59 that discuss the principles of risk assessment that were applied here is generally good. However, the point that the RfD and exposure values used were quite conservative (appropriately so) needs to be discussed so that the reader understands how protective the results are. The text starts to get at this issue by talking about bright lines, but does not go far enough.
5) There needs to be a fuller explanation of why the banned compound DEHP "dominates the calculation of the HI", as per L1627. I agree with the conclusion about the dominance, but explain why.
6) The Chapter 4 discussion provides a summary of what the CHAP did and the inherent limitations and uncertainties. It is good to have such a section. However, a few things should be considered.
    a) L1748ff discussed the criteria for selecting a study for use. Please double check whether these criteria were always applied. For example, did all the studies used follow EPA and OECD guidelines? Did the CHAP check that all the studies used by Kortenkamp and Faust met these criteria. L1758 says CHAP gave added weight to studies replicated in different labs. Indeed the CHAP mentioned this as an important point, but is "replicated" the correct word. It implies EXACT duplication of a test protocol. Or is reproduced a better term. L1760ff further complicates

understanding. It says that "...criteria to evaluate…and thereby derive reliable NOAELs... the final NOAELs used in the HI analysis are limited by the following." This language implies the many of the NOAELs are *not* reliable. I agree with the approaches taken. I have a problem with the language used to describe it.

b) L 1781 could use additional balance. Indeed, the reliance on data from one animal species can be limiting. However, some phthalates were tested in multiple species, and the rat was the most sensitive. Thus, the use of rat data are not a significant limitation, as implied.

7) L2093ff. This paragraph should be clarified. It says (L2096) that DBP exposures were possibly high enough to cause concern for human development or reproduction. Two sentences later it sates there was minimal concern for development effects. Meanwhile, this section is labeled "Reproductive."
8) The format used for Chapter 5 is excellent. Therefore, it should be followed. The exception to the format occurs when summarizing the NTP-CHRHR report. For example, under adverse effects, overall risk is summarized (e.g., L 2097ff, L2227, L 2354). Many of the risk statements say: "Typically, MOEs exceeding 100-1000 are considered adequate for public health; however, the cumulative risk of XXX with other anti-androgens should also be considered." I agree. However, there should be a discussion somewhere about how to deal with the "accumulation" of ever-larger "acceptable" MOEs to avoid an excessively conservative result. Also, most statements (e.g., 2212) say "However, CHAP recommends that U.S. agencies responsible for dealing with XXX exposures from food…with a view to supporting risk management steps." This presumes risk management steps are needed. Consider changing to saying "…with a view to supporting risk management evaluation."
9) L3261 raises a very important point, namely, that DIBP "today" does not represent a risk in toys because it is not used, but it is starting to be used. Hence, a recommendation to ban before it becomes a problem.
10) L3409 recommends a permanent ban on DHEXP. The text states that the database is weak, but that the toxicological profile is very similar to other antiandrogenic phthalates and DHEXP exposure would contribute to cumulative risk. I am not familiar with the studies on this or related phthalates. However, in my view, the text is not strong enough to support a ban. With such loose criteria, far more chemicals should be banned. One possibility would be to strengthen the arguments in the text. Another might be to state the concern, as does the current text, and call for an interim ban until more research is done. An interim ban is appropriately recommended for DIOP for database reasons (L3520).
11) L3471 recommends a permanent ban of DCHP. As stated above, the database appears weak so the rationale for a ban needs to be stronger. In addition, FDA has approved it, suggesting food safety.

12) L3877 recommends research on DEHA rather than any banning or non-banning action. However, the rationale appears inconsistent in that the MOE "from dietary DEHA exposure range from 770 to 290,000." Other phthalates (e.g., DINP) with larger MOEs are recommended for banning.
13) L4380 recommendation for TOTM. I agree, but this recommendation could be added to many others (i.e., get appropriate exposure info before use in toys and child care products.).
14) The report addresses phthalate substitutes as well as can be expected based on current knowledge. The report also correctly says that all future substitutes cannot be known now. Therefore, it would be useful to provide a general recommendation of what information should be required on new substitutes to enable an adequate risk assessment and the criteria that should be used to identify an interim or permanent ban for them.

2.6.1.3 Reviewer 3:

The overall conclusions are reasonable, but they are scattered throughout the report. It would be helpful to combine the highlights into separate conclusion and recommendation sections. Perhaps this is what is envisioned for the missing Executive Summary.

The CHAP provides the following general conclusions:

Section 2.4.1: Conclusions regarding the applicability of animal data to the evaluation of human exposure to phthalates

Section 2.4.2: Conclusions regarding neurodevelopmental studies

Section 2.5.5: Conclusions about ubiquitous phthalate exposure in the U.S. population, differences in exposure among various subpopulations, and correlations based on the HBM data

Section 2.6.6 and 2.6.7: Conclusions regarding the sources and amount of exposure to individual phthalates across different age group subpopulations based on the scenario-based exposure assessment

Section 2.7.1: Conclusion regarding appropriateness of assuming dose addition for phthalates

Section 5: Conclusions regarding toxicity data, exposure data, MOEs, and HIs (when calculated) for each phthalate and phthalate substitute

The CHAP also concludes that exposure estimates based on the HBM data and the exposure scenario-based modeled data are similar based on information in Table 2.14. There appear to be some notable differences (e.g., HBM data indicate a higher level of

DEHP exposure than the exposure scenario-based modeled data, and the reverse observation applies to DEP and, to a lesser extent, DINP). More importantly, Table 2-14 compares the average from the modeled data to medians from the HBM data sets. Why not present median results from the modeled data for a better comparison?

The overall recommendations are also reasonable; however, some should be clarified. The CHAP provides the following general recommendations:

Section 2.4.1: Recommends further study of human fetal exposure to phthalates results in adverse effects in addition to reduced anogenital distance.

Section 2.4.2: Recommends reduction in exposure to three phthalate metabolites based on concern about neurodevelopmental effects.

Section 5: Recommends instituting or eliminating bans for individual phthalates, and recommends further research for some individual phthalates.

Some recommendations in Section 5 are intended to prevent phthalates with the potential to cause adverse effects but that are not currently used from becoming a problem in the future. What's missing is a comprehensive discussion in one section of the report regarding the CHAP's recommendations for preventing future phthalate exposures of concern, including criteria used to decide that the level of concern is sufficient to warrant either a recommendation for further research and/or banning.

2.6.1.4 Reviewer 4:

Chapter 5 summarizes CHAP's recommendation on each phthalates using, presumably Weight-of-Evidence approach (see also earlier comments).

2.7 Research Needs

2.7.1 Question 6a: Are the CHAP's recommendations for future research appropriate?

2.7.1.1 Reviewer 1:

These were all fairly general and reasonable based on the dataset in hand which is very extensive and spans multiple decades of research and testing by academe, industry and government.

2.7.1.2 Reviewer 2:

1) The charge to CHAP does not specify identifying research needs. However, the CHAP is totally correct to add research needs. It is no surprise that the database on phthalates and substitutes are very deficient in many cases. Resources for research are always limited. Therefore, the CHAP should create a separate section on

research needs that identifies only the highest priorities. This can be difficult and time-consuming, but the CHAP is in a unique position to identify what are the big blocks of missing information.

2) At present, the organization of the document hides the research needs. For example, the main text of Chapter 2 has some needs sprinkled throughout (e.g., L542, L972, L2038). I do not disagree with the recommendations, but such needs could be attached to virtually every page. Also, research needs are expressed for some chemicals in Chapter 5. Please consider the following. Remove the research needs within Chapters 2, 3, and 4. Leave them as they are in Chapter 5 for specific chemicals. Then create a new research needs section that describes fundamental needs (e.g., extrapolating male reproductive MOA from rats to humans, understanding multiple pathways of exposure) and chemical-specific needs. The chemical-specific needs would be "copied" (and duplicate to) those in Chapter 5.
3) L972 is a recommendation about exposure reduction and research needed. Neither belongs here. The recommendations should be based on a synthesis of all the literature, not just the epidemiology literature.
4) L1024 is a recommendation based on the epi neurodevelopmental studies. First, it doesn't belong here. Secondly, it is far too strong based to the information discussed immediately above.
5) L 1803ff is a recommendation about interagency work and CPSC resources. I agree, however, it is misplaced here and should be moved to research needs.

2.7.1.3 Reviewer 3:

Some research recommendations are specific and appropriate (e.g., including phthalate substitutes in future biomonitoring). Others sound reasonable enough but are vaguely worded making it hard to understand the CHAP's priorities and the specific action being recommended. For example, in Section 5, the CHAP makes the following recommendation: "U.S. agencies responsible for dealing with DBP exposures from ***food, pharmaceuticals, and other products*** [emphasis added] conduct the necessary risk assessments with a view to supporting risk management steps." The same recommendation is made for some other phthalates except that different phthalate sources are listed. The CHAP lists "food, pharmaceuticals, and personal care products" for DEP, "food and other products" for BBP and DINP, "all sources" for DEHP, and "food and child care products" for DNOP and DIDP.

What specific action is the CHAP recommending for each phthalate and why is the recommendation applicable to the specific sources mentioned for each phthalate? The CHAP makes a similar recommendation for some phthalate substitutes, calling for the appropriate U.S. agencies to obtain exposure and toxicity data to assess risks.

The CHAP has dutifully and admirably responded to its charge, but I wonder if it is missing an opportunity to make broad recommendations regarding all of the stressors (chemical and non-chemical) that might adversely affect male development? A recent NRC committee charged with reviewing risk assessment practice recommended starting with the endpoint of concern (effects on male developmental effects) and identifying all factors contributing to this problem. CPSC's regulatory responsibilities are not that broad, so the CHAP is justified in not looking beyond the chemicals identified in its charge. However, it would be useful to note in the CHAP report any other chemical and non-chemical stressors that might adversely affect male development and whether they might co-occur with phthalates such that they could contribute to adverse effects in a cumulative manner. The CHAP has already done so to some extent with its sensitivity calculations showing how HIs would change with inclusion of non-phthalate stressors (Appendix D, Section 6).

2.7.1.4 Reviewer 4:

Yes. The development and application of physiologically-based pharmacokinetic (PBPK) model offers a sound approach to better understanding of total exposure from multiple sources. These models are often complex, but are increasingly feasible to build.

2.7.2 Question 7b: Are there any other suggestions for future work that would reduce the uncertainty in the risk assessment?

2.7.2.1 Reviewer 1:

Potential species differences in response to specific phthalates and phthalate mixtures for target lifestages remains an area of uncertainty that has not been completely resolved. Clearly there has been some discordance between rodent species, let alone primates and humans; although, it is not clear why mice appear less sensitive than rats, given that mice also exhibit adverse testicular and reproductive effects. In spite of the vast array of data available we still do not have precise mode of action information to explain the target(s) for toxicity or why certain periods of development are more sensitive than others. So for example, one can measure changes in rat fetal testosterone levels for specific phthalates during specific periods of gestation (e.g. GD 19-21). However even though phthalates can reduce this parameter at this time, it has very little consequence (if any) on adverse outcome in terms of the induction of malformations that constitute the phthalate syndrome. This is really why the human explant data is so suspect in that this has not been evaluated during the appropriate developmental stage is clearly very variable (given that a 50% reduction in fetal T is not statistically significant) and we are still not sure that such a reduction at this developmental age even if significant would be directly related to an adverse outcome. Indeed, one could argue that the limited human data both through fetal human testis implants in rodents

and the epidemiology findings would tend to indicate that the human is more like the rat than the mouse and not, as some have indicated the reverse.

2.7.2.2 Reviewer 2:

Yes, but I am not sufficiently expert to identify any beyond the need to get far more accurate information on exposure pathways from sources to people. Biomonitoring can be used to a degree in risk assessment, but it provides no help to risk management strategies.

2.7.2.3 Reviewer 3:

Nothing beyond recommendations listed in response to other charge questions.

2.7.2.4 Reviewer 4:

Decision and Science (NRC 2009) offers a view on uncertainty analysis as a way to promote risk assessment rather than hold it back because of gap in knowledge and assumption in the process. It also offers suggestions as to the balance of quantitative and qualitative characterization of uncertainty. It is important to understand uncertainty due to, for example, different models for ADI, or HI; it is also important to understand the aggregation of uncertainties from all stages of risk assessment and how the overarching uncertainty would form. This latter may deserve additional research efforts.

2.8 Other Comments

2.8.1 Question 8: Are there any other scientific issues or comments on the report?

2.8.1.1 Reviewer 1:

1) Role of PPARα in the production of the male reproductive effects (line 560). A recent paper (Hannas *et al.*, 2012) examined the ability of a number of phthalates and other agents to affect fetal testis testosterone production and gene expression related to sexual differentiation in the rat and used the prototypical PPARα- agonist Wyeth Wy-14,643 and while the expected gene changes did occur in the livers of dams, no comparable effects on fetal testosterone and gene expression comparable to that seen with a varied array of phthalates was observed. I believe this study finally eliminates some of the speculation on the role of PPARα in the induction of male reproductive malformations (and questioning the relevance to humans) via a predominantly antiandrogenic process as used by the Panel in this report.
2) Use of human fetal testis explant studies. I think I would also add to the comments on the Heger *et al*. (2012) study (line 2031), the recipient animal is an intact male rat with fully functioning testes and presumably normal adult levels of testosterone

in the circulation reaching the testis explant in the kidney, but no explanation as to how the presence of this high level of testosterone would impact the ability of any administered phthalates to reduce testis explant androgen levels and any associated genes is provided. This is totally different from the situation in a pregnant rat dam with levels of T originating in the fetal testis only and no supplementation from the dam (in this case the host!).

3) Thalidomide (line 700). This comment is really not true. Developmental toxicity was seen in other species. However, phocomelia, the defect observed in humans was not seen in a number of the animal species (rodents) but was seen in rabbits.
4) Use of NOAEL (line 749). Just because a study does not have a NOAEL does not make it unsuitable for analysis. Selection of dose levels to provide a measure of the shape of the dose response curve could be considered very useful than a poorly powered study with a clear NOAEL or dose spacing that did not provide information close to a POD. In the case of a study without a NOAEL, but a reasonable dose response for the selected toxicity, then apply the BMD approach for further risk assessment.
5) Table 2.1. Why not use the Blystone *et al*. 2010 paper for DEHP – it has the most rigorous dose response analysis of the phthalate syndrome with the same NOAEL as that used in the table.
6) Line 822. Not sure how an OECD single generation study really provides any information on the phthalate syndrome, which would be a very different situation from a multigeneration study – where the syndrome was first noticed!
7) Line 951. The recent paper (Dean and Sharpe, 2013) also confirms the utility of AGD measurements in humans as a surrogate for testosterone status.
8) Human biomonitoring data. Line 1073. Did the Panel give any thought to looking at other (smaller studies) that may provide some indication of internal dose (and particularly to fetuses) rather than back calculating to an administered dose level from a urine measurement? There is a concern that a comparison of internal doses may be a more useful comparator between animal effect and humans and that administered doses (on a mg/kg/d basis) may be underestimating fetal exposure to phthalate monoesters (see comments above) especially in comparing human exposures with animal effects at specific dose levels of agents.
9) Line 1129. How have different PK parameters for individual phthalates been taken into account in the confidence ascribed to spot urine samples?
10) Have not pentyl phthalate metabolites also been observed in human urine samples? I thought CDC had made attempts at metabolite measurements and detected it in human samples. It would be interesting to see how this potent phthalate would also influence mixture and other conclusions in the 2.6.7 section and 5.4.4.4.
11) Table 2.7. While I understand the process undertaken here in moving from human urine values back to intake data, it seems we are missing something more about

internal dose (e.g., blood or tissue concentrations). There are physiologically-based pharmacokinetic (PBPK) models out there that might aid this. The concern is more on the extrapolation issues between animal and human data and these comparisons are always more compelling based on an internal dose estimates, than an intake level. It is not unusual to see a much larger intake dose in rodents equate to a much lower intake dose in humans when blood or critical tissue concentrations are comparable. I would have thought the limited amniotic fluid metabolite measurements between rat and human would have provided useful information on potential inter-species extrapolation when comparing internal dose (see also section 4.2). Moreover, as mentioned above, this is the most direct measurement of fetal exposure – the critical exposure group for phthalates. Similarly, although exposure to pregnant women has to occur for fetal exposure to ensue (see Table 2.8) have the estimations been performed on what this might mean for AF and fetal exposure from these urine levels (even if at the appropriate time for human development of the reproductive system)?

12) Case 3. Line 1594. Could the Panel provide some further explanation of the selection of an uncertainty factor of 100 in this case? Likewise in Table 2.15 some footnote on selection of why different uncertainty factors were chosen would be helpful so that the Table "stands alone".
13) Table 2.16. I assume that 1, 2 and 3 in the table are the different case studies? Could be more clearly annotated.
14) Line 1908. I thought this section could also benefit from some statements on the variability of kinetics due to mixture exposures. Since diesters are competing for similar enzymes irrespective of their potential activity *in utero*, how does a higher concentration in tissues (derived from urine levels) of an inactive ester impact the metabolism and kinetics of an active phthalate in the same mixture. Are levels so low that there would be no impact? Are there scenarios where half life may be changed due to poorer absorption, or clearance, or both?

2.8.1.2 Reviewer 2:

1) General: As is obvious from my detailed comments, I am very supportive of the risk assessment and risk management conclusions. However, the evidence to support those conclusions is strewn about the document and sometimes missing. This *greatly* weakens the case when a non-expert or scientific critic reviews the document. This can be remedied by asking a highly qualified technical editor take maybe 75% of the material out of Chapter 2 (Background and Strategy) and move it to Chapters 3 and 4. Right now, far too many unreferenced "facts" and "conclusions" are in Chapter 2 and often they repeat what is said or substantiated better in chapters 3 and 4. Specific comments to support this general statement are provided under responses to individual charge questions.

2) Communication elements. The main body of the report needs to relate better to the technical/policy expertise of the audience. I suggest adding a new section to Chapter 2 that describes the risk assessment process in no more than two paragraphs. This allows for definition of terms (e.g., uncertainty factor, RfD) and principles (e.g., MOA, WOE, MOE, dose in an animal must be extrapolated to humans). The reader should get a clear understanding of how protective the risk components are. Examples:
    a) P4 footnote refers to "daily intake" and "reference dose". Most scientists and policy makers don't know what an RfD is.
    b) Terms MUST be defined. It is especially important to define dose, intake, exposure, hazard, risk.
    c) P16 L802ff. This is generally a tutorial about the types of endpoints studied. This is a good example of the difference in level of detail. For example, why describe general toxicological studies but not describe an RfD or uncertainty factor?
3) Criteria for inclusion of references in the document are stated, but in some cases, it is not certain whether they were adhered to. I recall that one of the main elements of the Information Quality Act that applies to regulatory agencies' dissemination of information is a requirement that the information be reproducible (able to be reproduced because enough information was presented in the study; not a requirement that it has been reproduced, although that would be good to know). There should be a concerted attempt to ensure that all the criteria cited in CPSC's policies for implementation of the Information Quality Act were followed. It is clear that information in a peer-reviewed journal would be acceptable, but information in an abstract or a presentation would not. Examples of some issues:
    a) L339 "…all studies available in the public domain were analyzed." Is the EPA/CPSC database you used "in the public domain?"
    b) L340 describes how CHAP assessed the quality of the report. This cites an OECD document that most have not read, so it is not clear how the information was assessed. Please be more specific using common criteria like: in the peer-reviewed literature (not just journals; some books and proceedings are peer-reviewed); in EPA/CPSC databases from manufacturers and then describe whether the federal organization reviewed the reports for quality; in the gray literature and then describe whether CHAP or its contractors actually read the papers. This issue comes up again on L1755ff. Again, the document seems to counter-state things. For example, L1756 says only OECD protocol studies should be used, but L1760 says other studies were used. I agree that OECD protocols should not limit the use of data; my concern is over the written discrepancy.

c) Later (L625) more specifics are given, but the criteria should all be in one place (maybe even a specific subsection). Since there is often mistrust of industry-generated data, the extent to which the industry followed good laboratory practice (GLP) protocols and the regulatory agency reviewed that data should be described.

d) L500 refers to a study "…reported only in abstract form, Marsman (1995)…" I strongly recommend deleting ALL abstracts from the document. This case is especially a problem because one may surmise there was something very wrong if the author has not published work from 17 years ago. Also, how can the quality of any abstract be assessed according to CHAP criteria?

e) L527 refers to "unpublished studies" from a presentation. Such data do not meet CHAP criteria and should be deleted.

f) L630ff says that some unpublished data will be used. It is important to expand on this. For example, did the CHAP have access to the materials, methods, and results and evaluate them for quality. If the CHAP used a recent abstract with the understanding that the work will be published very soon, it is important to check for the full reference before this document is published. This whole topic is too important to rely on work that cannot be reproduced because of inadequate information.

g) L677-679 is of *substantial* concern. This sentence essentially says that more reliance was put on positive studies of quality and implies that negative studies of quality were of lesser value. I am confident that the authors carefully considered *all* papers that had sufficient scientific quality.

h) L1099 says that HBM data from SFF, which are crucial to the risk assessment, were "provided to the CHAP by Dr. Shanna Swan and are published in part in…" What does "in part" mean. If this is a case of needing more data from the study than were published, I understand. However, in the interest of transparency, these unpublished data should be made publically available.

4) P 9 L607ff Section 2.3.1 Use of Animal Data.

a) All risk terms need to be defined, as stated above

b) The greatest concern is that the discussion does not deal with interspecies extrapolation from toxicokinetic and toxicodynamic perspectives. Given the great interspecies differences and the exceeding high doses required to cause effects in rats, such a discussion is needed.

c) L666 is very unclear.

d) L673. I disagree with this statement that OECD protocols are "most useful for risk assessment. I agree that they are useful, but not "most." I am not

familiar with the details of these protocols, but most international or government-approved protocols are quite simplistic and need to be buttressed with research (as opposed to standardized testing) for good risk assessments. It appears that the authors would agree since they state the value of high quality university-based research.

e) L676 This says …"experimental design in the context of standard protocols." Experimental design has a science to it that extends beyond that of standard protocols.

f) L677 This says "route of exposure". This needs to be expanded to include dose-rate, dose, dosing regimens (e.g., windows of vulnerability).

g) L690 This sentence requires more than animal toxicology can ever do, namely "definitely predict."

h) The issue of maternal toxicity should be discussed because fetal effects in rats are major bases of the conclusions and very high doses were used.

5) P18 L926ff. Epidemiology
    a) The first paragraph should offer a very brief discussion of the difference between association and causality.
6) L1508ff and L1587ff which discuss POD do not appear to agree. P 51 criticizes the POD and says HI used, whereas p 52 uses the POD.
7) L2164 In several places, the text says that it was difficult to ascertain how many rats were used in the Kim *et al*. 2010 study. So, I took a look at it. On several figures and tables, the number of rats is indicated. In many of the results tables, the data are given in terms of pups. They also did a statistical analysis. I have not personally done such studies, but can't the adequacy of the n be determined from the information provided?
8) Miscellaneous minor or editorial comments
    a) Several tables do not stand alone or are either not immediately clear or confusing. All should be examined for editing. As examples:
        1. L 667 Table 2.16. This table does not stand alone. For example, each box has three numbers; I had to ask what these were (answer median, 95^{th}, and 99^{th} percentile). Also the "% with HI>1". % of what. What percentile. Also, please check the HI row, second box. What is the "S"?
        2. Table 2.15 does not stand alone. Where are the references?
        3. Table D-1 (p14 Appendix D) doesn't stand alone. What do the bars represent; what does the box represent? Appendix D describes the derivation of the HI.
        4. Table D-8 (L568ff) says that Case 1 was "altered from Kortenkamp and Faust 2010". I checked the reference and didn't see any

alterations. Also, earlier the document does not discuss any "alterations".

b) Table 2.16 needs substantial editing. It is virtually impossible to look at the table and determine what the three values in each box are. The Label above DIBP, DBP, etc. says "RfD Case". The bottom left box say % but not % of what.
c) L749 the parentheses says NOAEL, but a LOAEL is described.
d) L812 shouldn't have a dot
e) L1575 the word "animals" should be inserted at the end to ensure people don't think you are referring to humans.
f) L1594 refers to Table 2.2 which is on p15. It has no RfDs.
g) L1896ff. It would be useful to add the phthalate t1/2 here or in a table.
h) L2231 Please indicate the animal species used.
i) L2637 Please clarify that this means the data are not relevant to humans.
j) L2688 Please clarify the location of "Figure 6."
k) L3508 apparently has an author note "check reference". I state it so this note doesn't end up in the final version.
l) L1663. This table has DnBP, but elsewhere this chemical is abbreviated DBP.
m) L1709 misspelling

2.8.1.3 Reviewer 3:

On behalf of the CPSC, the CHAP has undertaken a substantial and well-crafted assessment of cumulative risk from the general U.S. population's exposure to phthalates. All comments and recommendations provided in this review are intended to make sure that the CHAP's work is interpreted correctly and used wisely.

To summarize, the CHAP's report could provide even greater value if it clearly provided conclusions and recommendations for:

- Optimizing biomonitoring data collection to support future cumulative risk assessments (e.g., chemicals of interest included and samples collected in a manner that reflects population variability over the exposure period of interest given possible adverse effects).
- Gathering additional and better product use and exposure pattern data in a systematic manner to support future exposure scenario-based assessments, which are needed to identify and reduce sources of exposure as needed and to estimate future exposures that cannot be measured with HBM data.
- Checking the applicability of rat toxicity data to humans and the assumption of dose additivity (i.e., dose the CHAP think that the unpublished data it reviewed are sufficiently persuasive to justify further study?).

- Checking whether there are any exceptions to the assumption of dose additivity (i.e., on page 18, lines 895-897, the CHAP refers to a finding from Christiansen et al. [2009] suggesting synergy among DEHP, vinclozolin, finasteride, and prochloraz on malformations of external sex organs).
- Developing a strategy for the CPSC to prevent or, where they have already occurred, reduce problematic exposures. The CPSC's regulatory responsibilities do not encompass all sources of exposure to phthalates and other stressors that might adversely affect male development. However, it would be useful to acknowledge this broader perspective in the CHAP's recommendations and discuss how to protect fetuses and young children within our current system of fragmented regulatory oversight.

Editorial Suggestions

1) Page 8, line 528 and 538: Appears that "rat" should be replaced with "mice"
2) P12, lines 705-707: "Data from human studies of reasonable quality generally are a stronger signal of risk to humans than findings in animal studies. However, in the absence of other data, findings in animals should be assumed to be relevant for prediction of risk to humans." Some qualification is needed here because there might be cases where there is no reason to believe that the animal data are relevant to humans.
3) The CHAP refers to: "…the traditional approach to risk assessment with its focus on single chemicals one-by-one may inadequately address the health risks that might arise from combined exposures to multiple chemicals" (Page 18, lines 928-930). While much risk assessment is performed as the authors indicate (e.g., most drinking water quality standards), many risk assessments endeavor to account for cumulative risk. This has been done in the Superfund program for decades and more recently in the regulation of pesticides (See the 2008 NRC report regarding phthalates for more examples). There is nothing to be gained by not acknowledging some successful precedent for cumulative risk evaluations.
4) Page 26, lines 1132-1134: "rather similar" What is meant by this phrase exactly and what is its value to the assessment? Similar with respect to the potential for toxicity? For example, exposures that appear "rather similar" might be significantly different with respect to their potential to cause adverse effects. Plus some look "rather different" to me such as DEHP metabolites in infants as reported by CHAP based on SFF data (Table 2.5, last row of first page)
5) There is no discussion or reference to Table 2.10 in the text. This table is taken from Appendix E-1 (Table E1-2) where some explanation is provided. I suggest this

explanation, or some portion of it, be moved to the main body of the report to improve clarity.

6) Pages 58-59, lines 1712-1725: the CHAP advocates for use of MOEs, expressing concern about RfDs, which might be perceived as "bright lines" that are fictional given the uncertainty underlying them. The CHAP explains (page 58, lines 1717-1720): "In taking this [MOE] approach, it was possible to avoid misunderstandings that might have occurred had CHAP used points of departure and combined them with uncertainty factors to arrive at "tolerable exposures" or reference doses." The concern is valid. However, the CHAP did in fact calculate RfDs, so it is confusing to suggest that they did not. Moreover, the CHAP calculated HIs based on a range of RfDs for each antiandrogenic phthalate, thus clearly communicating at least some of the uncertainty associated with quantifying a "safe" dose. Therefore, it is not apparent that the MOE approach is superior to the HI approach, and the MOE approach does not allow for an assessment of cumulative risk as the HI approach does. The report would benefit from the CHAP clarifying its views on and use of HIs and MOEs in formulating its opinions and recommendations.

7) MOEs for non-antiandrogenic phthalates and phthalate alternatives are discussed in the context of making recommendations, but it would be helpful to tabulate them in the report.

8) Appendix C, consider adding a new reference: Mouritsen A, Frederiksen H, Sørensen K, Aksglaede L, Hagen C, Skakkebaek N, Main K, Andersson A, Juul A. (2013). Urinary Phthalates from 168 Girls and Boys measured twice a year during a 5 Year Period: Associations with Adrenal Androgen Levels and Puberty. J Clin Endocrinol Metab. doi:10.1210/jc.2013-1284.

9) In Appendix D, the CHAP refers to three antiandrogenic compounds, BPA, PPB, and BPB (i.e., bisphenol A and two parabens). It would be helpful to define these abbreviations in this appendix.

10) In Appendix E1, equation number 2 is missing a "surface area" variable with units of cm^2. However, the exposure calculations are correct based on review of the excel spreadsheet provided with the review package (i.e., with the assumption of a 10 cm^2 exposure area). While references to evolving literature must cease at some point in time, there are a few new ones that would be useful to reference in the final version of this report:

    a) Koch HM, Lorber M, Christensen KL, Pälmke C, Koslitz S, Brüning T. (2013). Identifying sources of phthalate exposure with human biomonitoring: Results of a 48h fasting study with urine collection and personal activity patterns. Int J Hyg Environ Health. doi:pii: S1438-4639(12)00138-1.10.1016/

b) Langer S, Bekö G, Weschler CJ, Brive LM, Toftum J, Callesen M, Clausen G. (2013). Phthalate metabolites in urine samples from Danish children and correlations with phthalates in dust samples from their homes and daycare centers. Int J Hyg Environ Health. doi:pii: S1438-4639(13)00052-7.10.1016/

c) Hernández-Díaz S, Su YC, Mitchell AA, Kelley KE, Calafat AM, Hauser R. (2013). Medications as a Potential Source of Exposure to Phthalates among Women of Childbearing Age. Reprod Toxicol. (37):1–5.

d) Parlett LE, Calafat AM, Swan SH. (2013). Women's exposure to phthalates in relation to use of personal care products. J Expo Sci Environ Epidemiol. (23):197–206.

2.8.1.4 Reviewer 4:

The Chronic Hazard Advisory Panel (CHAP) conducted a review to examine the health risk associated with phthalates, including phthalates and phthalate alternatives in children's toys and child articles. The findings and recommendations are presented in the Peer Review Draft (the draft). The review is a huge undertaking about risk assessment of cumulative/total exposure to phthalates with an emphasis on vulnerable subgroups who are more likely to be also exposed through toys and personal care products. Therefore a part of my comments is about the risk assessment approach that the draft took.

Systematic Review The draft includes a comprehensive review of the literature concerning health hazard and risk of phthalates. The appendices A, B, and C are a review on developmental toxicity, reproductive and other toxicity, and epidemiological studies, respectively. For each phthalate, the draft reviews available studies, evaluates the quality of the study and appropriateness of the resulting database, and synthesize all evidence. This part involves hazard identification as well as dose-response assessment of a risk assessment process (NRC 1983; EPA, 2005; NRC 2009). However, the draft provides little description of CHAP's literature search strategy with respect to the scope (e.g. type of databases, years of publications). Thus it is difficult to judge whether the search is inclusive. For example it is mentioned several times in Appendix B that search was done in PubMed. Is PubMed the only database searched and is it sufficient?
In Chapter 5 of the draft, CHAP evaluates the quality of each study identified regarding number of dose groups, number of animals per group, exposure time window, for instance. These criteria are relevant about dose-response and statistical power, but there are other criteria to determine if a study should be included in a risk assessment.
The field of risk assessment is moving quickly to adopt the approach of *systematic review* as a guiding principle as well as a practical process for literature review, evidence gathering and synthesis in support of risk assessment (EPA 2004; NRC 2011). As a gold standard, systematic review enhances objectivity, transparency, as well as

credibility. The lack of description of the search scope, inclusion and exclusion criteria makes it difficult to evaluate to what extend the current review is thorough and complete in identifying the latest and best scientific evidence.

Weight of Evidence. In Chapter 5 of the draft CHAP makes recommendation on each phthalate under the consideration. It lists the following set of criteria (P67-68):

1) Nature of the adverse effects reported in animal and human study. Is there evidence of reproductive and developmental effects (e.g. Phthalate Syndrome)
2) Relevance of findings from animal studies to human
3) Weight of evidence, e.g. appropriate study design, power, confounders, replications in other studies
4) Risk to human: likely exposure to human, hazard, dose-response, POD, exposure relative to POD (e.g. MOE, HI)
5) Recommendation narrative
6) How a recommendation would affect children

It is unclear why a recommendation descriptor (5) is a criterion of making the recommendation. "Weight of evidence" (3) is about the quality and quantity of the available database based on the design and statistical power of studies. So taken together, the entire section for each phthalate is in fact a weight of evidence narrative. Weight of evidence is a process of collective evaluation of all pertinent information so that the full impact of biological plausibility and coherence is adequately considered. Judgment on the weight of evidence involves consideration of the quality and adequacy of data and consistency of responses induced by the chemical stressor, but also requires combined input of relevant disciplines: toxicology, biology, chemistry, epidemiology, statistics, etc. (EPA, 2004). For example, mode of action and biological plausibility are a part of weight of evidence evaluation. Following EPA's cancer guidelines (EPA 2005), a weight of evidence narrative can generally includes:

- conclusions about human toxicity potential
- a summary of the key evidence supporting these conclusions, including information on the type(s) of data
- epidemiologic or experimental conditions that characterize the adverse health effects
- a summary of potential modes of action and how they reinforce the conclusions,
- indications of any susceptible populations or life stages, when available, and
- a summary of the key default options invoked when the available information is inconclusive

Recommendation. Whether or not there are reproductive and developmental effects appears to be a driving factor in CHAP's recommendation for each phthalate. This

needs to be made transparent in the weight of evidence process. For instance the permanent ban on DNOP is recommended to be lifted. There is little evidence of reproductive and development effects of DNOP, but evidence on systemic effects is strong. With respect to population exposure, the MOE is higher. In comparison, because of the evidence of reproductive and developmental effects, DIBP and DINP made to the permanent ban list despite a higher MOE (Table 2.17) and less than 1% median contribution to the overall HI of 5 phthalates. The currently unbanned phthalates DPENP and DHEXP also made to the permanent ban list for the reproductive and developmental effects even though there is little information about their MOE or HI. Without a clear description of the weight-of-evidence process that is inherent in the CHAP recommendation, the treatment of different phthalates seems uneven because the CHAP did not calculate MOEs of DNOP in relation to systemic effects. In contrast, CHAP identified the liver as a target organ for DIBP, calculated a number of reference NOAELs based on systemic effects (in the liver or kidney).

HI is a useful concept in assessing the potential risk at the current population exposure level. If HI is used at all in assessing likely exposure level relative to an established POD, it is often not made clear, but should be made more transparent.
The presentation of recommendation is a summary of the risk assessment process of each phthalate. But in the process key information may have been omitted. For instance, many of the references were omitted in the recommendation section (e.g. L 2980-2981; L 3239-3231) and the reader has to go back the relevant appendices.
P90, L3314-3318: what evidence and rationale led to the conclusion that "DPENP is clearly among the most potent phthalates regarding developmental effects". Given that there are only two studies on rats and no human evidence available, as reported by CHAP, the process leading to this conclusion is not transparent. The recommendation needs much more support and articulation. In contrast to the discussion on other phthalates, there is no discussion in the "Risk" section on MOE or any POD.

Editorial comments:

- The term "sufficient design" (e.g. L3836-38) or "sufficient number" (e.g. L3981) is misleading. A study design can be sufficient in statistical power to detect dose effect or determining a POD, or to confirm a biological effect. A generic "sufficiency" conveys a false sense of security.

- The report can benefit from a careful proof reading.

.

3 References

Barlow NJ, Foster PMD. (2003). Pathogenesis of Male Reproductive Tract Lesions from Gestation Through Adulthood Following in Utero Exposure to Di(n-butyl) Phthalate. Toxicol Pathol. 31(4):397-410.

Blystone CR, Kissling GE, Bishop JB, Chapin RE, Wolfe GW, Foster PM. (2010). Determination of the di(2-ethylhexyl) phthalate NOAEL for reproductive development in the rat: importance of the retention of extra animals to adulthood. Toxicol Sci. 116(2):640-646.

Calafat AM, McKee RH. (2006). Integrating biomonitoring exposure data into the risk assessment process; Phthalates [diethyl phthalate and di(2-ethylhexyl) phthalate] as a case study. Environ Health Perspect. 114(11):1783-1789.

Calafat AM, Brock JW, Silva MJ, Gray LE, Jr., Reidy JA, Barr DB, Needham LL. (2006). Urinary and amniotic fluid levels of phthalate monoesters in rats after the oral administration of di(2-ethylhexyl) phthalate and di-n-butyl phthalate. Toxicology. 217(1):22-30.

Cook JC, Klinefelter GR, Hardisty JF, Sharpe RM, Foster PM. (1999). Rodent Leydig cell tumorigenesis: a review of the physiology, pathology, mechanisms, and relevance to humans. Crit Rev Toxicol. 29(2):169-261.

David RM, Moore MR, Finney DC, Guest D. (2000). Chronic toxicity of di(2-ethylhexyl)phthalate in rats. Toxicol Sci. 55(2):433-443.

Dean A, Sharpe RM. (2013). Anogenital distance or digit length ratio as measures of fetal androgen exposure: relationship to male reproductive development and its disorders. J Clin Endocrinol Metab. 98(6):2230-2238.

Foster PMD. (2007). Induction of Leydig Cell Tumors by Xenobiotics. In: Payne A, Hardy M. (Eds.) *The Leydig Cell in Health and Disease*. Human Press, Totowa NJ. pp. 383-392.

Frederiksen H, Kranich SK, Jørgensen N, Taboureau O, Petersen JH, Andersson AM. (2013). Temporal variability in urinary phthalate metabolite excretion based on spot, morning, and 24-hour urine samples: considerations for epidemiological studies. Environ Sci Technol. (47):958−967.

Gaido KW, Hensley JB, Liu D, Wallace DG, Borghoff S, Johnson KJ, Hall SJ, Boekelheide K.

(2007). Fetal mouse phthalate exposure shows that gonocyte multinucleation is not associated with decreased testicular testosterone. Toxicol Sci. 97(2):491-503.

Gray TJB, Rowland IR, Foster PMD, Gangolli SD. (1982). Species differences in the testicular toxicity of phthalate esters. Toxicol Lett. 11:141-147.

Hannas BR, Lambright CS, Furr J, Evans N, Foster PM, Gray EL, Wilson VS. (2012). Genomic biomarkers of phthalate-induced male reproductive developmental toxicity: a targeted RT-PCR array approach for defining relative potency. Toxicol Sci. 125(2):544-557.

Heger NE, Hall SJ, Sandrof MA, McDonnell EV, Hensley JB, McDowell EN, Martin KA, Gaido KW, Johnson KJ, Boekelheide K. (2012). Human fetal testis xenografts are resistant to phthalate-induced endocrine disruption. Environ Health Perspect. 120(8):1137-1143.

Hernández-Díaz S, Su YC, Mitchell AA, Kelley KE, Calafat AM, Hauser R. (2013). Medications as a Potential Source of Exposure to Phthalates among Women of Childbearing Age. Reprod Toxicol. (37): 1–5.

Holm M, Rajpert-De Meyts E, Andersson AM, Skakkebaek NE. (2003). Leydig cell micronodules are a common finding in testicular biopsies from men with impaired spermatogenesis and are associated with decreased testosterone/LH ratio. J Pathol. 199(3):378-386.

IARC (International Agency for Research on Cancer) (2000). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Volume 77. Some Industrial Chemicals. World Health Organization.

Ito Y, Yamanoshita O, Asaeda N, Tagawa Y, Lee CH, Aoyama T, Ichihara G, Furuhashi K, Kamijima M, Gonzalez FJ, Nakajima T. (2007). Di(2-ethylhexyl)phthalate induces hepatic tumorigenesis through a peroxisome proliferator-activated receptor alpha-independent pathway. J Occup Health. 49(3):172-182.

Kim TS, Jung KK, Kim SS, Kang IH, Baek JH, Nam H-S, Hong S-K, Lee BM, Hong JT, Oh KW, Kim HS, Han SY, Kang TS. (2010). Effects of *in utero* exposure to di(*n*-butyl) phthalate on development of male reproductive tracts in Sprague-Dawley rats. J Toxicol Environ Health A. 73:1544-1559.

Koch HM, Rossbach B, Drexler H, Angerer J. (2003). Internal exposure of the general population to DEHP and other phthalates - determination of secondary and primary phthalate monoester metabolites in urine. Environ Res. 93:177-185.

Koch HM, Lorber M, Christensen KL, Pälmke C, Koslitz S, Brüning T. (2013). Identifying sources of phthalate exposure with human biomonitoring: Results of a 48h fasting study with urine collection and personal activity patterns. Int J Hyg Environ Health. doi:pii: S1438-4639(12)00138-1.10.1016/

Kortekamp A, Faust M. (2010). Combined exposures to anti-androgenic chemicals: steps towards cumulative risk assessment. Int J Androl. 33:463-474.

Langer S, Bekö G, Weschler CJ, Brive LM, Toftum J, Callesen M, Clausen G. (2013). Phthalate metabolites in urine samples from Danish children and correlations with phthalates in dust samples from their homes and daycare centers. Int J Hyg Environ Health. doi:pii: S1438-4639(13)00052-7.10.1016/

Marsman D. (1995). NTP technical report on the toxicity studies of dibutyl phthalate (CAS No. 84-74-2) administered in feed to F344/N rats and B6C3F1 mice. Toxic Rep Ser. 30:1-G5.

Mouritsen A, Frederiksen H, Sørensen K, Aksglaede L, Hagen C, Skakkebaek N, Main K, Andersson A, Juul A. (2013). Urinary Phthalates from 168 Girls and Boys measured twice a year during a 5 Year Period: Associations with Adrenal Androgen Levels and Puberty. J Clin Endocrinol Metab. doi:10.1210/jc.2013-1284.

Mylchreest E, Sar M, Cattley RC, Foster PMD. (1999). Disruption of Androgen-Regulated Male Reproductive Development by Di(n- Butyl) Phthalate during Late Gestation in Rats Is Different from Flutamide. Toxicol Appl Pharmacol. 156(2):81-95.

NRC (National Research Council) (1983). Risk Assessment in the federal government: Managing the process. National Academies Press: Washington, D.C.

NRC (National Research Council) (2006). Human Biomonitoring for Environmental Chemicals. Committee on Human Biomonitoring for Environmental Toxicants, National Research Council. Washington, D.C. http://www.nap.edu/catalog.php?record_id=11700.

NRC (National Research Council) (2008). Phthalates and Cumulative Risk Assessment. The Tasks Ahead. Committee on the Health Risks of Phthalates, National Research Council. Washington, D.C.

NRC (National Research Council) (2009). Science and Decisions: Advancing Risk Assessment. Committee on Improving Risk Analysis Approaches Used by the U.S. EPA, National Research Council. Washington, D.C.

NRC (National Research Council) (2011). Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. National Academies Press, Washington, D.C.

Parlett LE, Calafat AM, Swan SH. (2013). Women's exposure to phthalates in relation to use of personal care products. J Expo Sci Environ Epidemiol. 23:197–206.

Saravanabhaven G, Guay M, Langlois E, Giroux S, Murray J, Haines D. (2013). Biomonitoring of phthalate metabolites in the Canadian population through the Canadian Health Measures Survey (2007-2009). Int J Hyg Environ Health. pii: S1438-4639(12)00145-9. doi: 10.1016/j.ijheh.2012.12.009. [Epub ahead of print]

Silva MJ, Reidy JA, Herbert AR, Preau JL, Jr., Needham LL, Calafat AM. (2004). Detection of phthalate metabolites in human amniotic fluid. Bull Environ Contam Toxicol. 72(6):1226-1231.

USEPA (U.S. Environmental Protection Agency) (2000). Supplementary Guidance for Conducting Health Risk Assessment of Chemical Mixtures. EPA/630/R-00/002

USEPA (U.S. Environmental Protection Agency) (2004).An Examination of EPA Risk Assessment Principles and Practices. EPA/100/B-04/001

USEPA (U.S. Environmental Protection Agency) (2005). Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001b NCEA-F-0644b Risk Assessment

Voss C, Zerban H, Bannasch P, Berger MR. (2005). Lifelong exposure to di-(2-ethylhexyl)-phthalate induces tumors in liver and testes of Sprague-Dawley rats. Toxicology. 206(3):359-371.

Wittassek M, Koch HM, Angerer J, Brüning T. (2011). Assessing exposure to phthalates - The human biomonitoring approach. Mol Nutr Food Res. 55:7-31.

Appendix A

Peer Review Instructions and Charge Questions

This page left intentionally blank.

Peer Review Instructions

Dear Panel Members,

Thank you again for agreeing to serve as a peer reviewer of the CHAP Phthalates Draft Report. We greatly appreciate your participating in this review. This email provides the review materials and instructions.

Document and Charge Questions - Attached you will find the draft document for your review, along with a Word file of the charge questions. Please address each charge question by adding your answers to the Word document. Please provide clear rationales and support for your opinions and indicate page and paragraph where applicable.

Confidentiality Note – Please remember that this draft report is confidential and should not be shared with anyone else or cited or referenced. We will ask you to delete and destroy all copies of the text upon completion of the review.

References – We will mail you a DVD of the references. If you need a CD format, please let Ann Parker (parker@tera.org) know. This disk and the references it contains should not be copied or used for any other project. We'll ask you to destroy this disc upon completion of the project.

Due Date - Your written review should be **returned to Jacqueline Patterson (patterson@tera.org) by email no later than July 19, 2013**. We will then compile a single draft report of all the reviewers' comments organized by charge question. We will forward this draft report to you so that you may review it and revise your comments if you feel that is needed.

Questions - If you have questions regarding this review, please do not hesitate to call or email me. Thank you again for being willing to do this review. TERA and the CPSC very much appreciate your assistance.

Please confirm via email that you have received the review materials.

Charge for Peer Reviewers

Peer Review of the CHAP Draft Report on Phthalates and Phthalate Substitutes

Background

On August 14, 2008, President Bush signed into law the Consumer Product Safety Improvement Act of 2008 (PL 110-314) (CPSIA), which, among other things, required the U.S. Consumer Product Safety Commission (CPSC) to appoint a Chronic Hazard Advisory Panel (CHAP) to examine the potential health effects of phthalates in children. According to the requirements in Section 28 of the Consumer Product Safety Act (15 U.S.C. § 2077), the Commission appointed seven CHAP members from a list of scientists nominated by the President of the National Academy of Sciences. The CHAP met for the first time in April 2010, and it must complete its final report by April 2012.

The task of the CHAP is to conduct a *de novo* examination of the risks associated with phthalates and phthalate alternatives in children's toys and child care articles. Specifically, Section 108 (b)(2)(B)(i)-(viii) of the Act states that the panel will:

(i) examine all of the potential health effects (including endocrine disrupting effects) of the full range of phthalates;

(ii) consider the potential health effects of each of these phthalates both in isolation and in combination with other phthalates;

(iii) examine the likely levels of children's, pregnant women's, and others' exposure to phthalates, based on a reasonable estimation of normal and foreseeable use and abuse of such products;

(iv) consider the cumulative effect of total exposure to phthalates, both from children's products and from other sources, such as personal care products;

(v) review all relevant data, including the most recent, best-available, peer-reviewed, scientific studies of these phthalates and phthalate alternatives that employ objective data collection practices or employ other objective methods;

(vi) consider the health effects of phthalates, not only from ingestion, but also as a result of dermal, hand-to-mouth, or other exposure;

(vii) consider the level at which there is a reasonable certainty of no harm to children, pregnant women, or other susceptible individuals and their offspring, considering the best available science, and using sufficient safety factors to account for uncertainties regarding exposure and susceptibility of children, pregnant women, and other potentially susceptible individuals; and

(viii) consider possible similar health effects of phthalate alternatives used in children's toys and child care articles.

Charge Questions

The CHAP has requested a broad, high-level peer review of the report. That is, the CHAP is primarily interested in a review of the overall risk assessment process that they applied to phthalates. In particular, they requested review of those areas of the risk assessment process that employ novel methodologies, such as the development of distributions of hazard indices as a means to assess risk from multiple phthalates (question 2). However, it is not the intention of the CHAP to dissuade the peer reviewers to comment on any aspect of the report that they deem significant.

1. **Analysis of Biomonitoring Data**
    a. Is the CHAP's analysis of biomonitoring data appropriate for assessing cumulative risk?
    b. Is the use of spot urine samples appropriate for estimating population exposure?
    c. Is the use of spot urine samples likely to underestimate or overestimate the median or upper bound exposures?
    d. Does the report adequately characterize the uncertainty of the biomonitoring data and approach?

2. **Cumulative Risk Assessment**
    The CHAP calculated hazard indices for individuals exposed to multiple phthalates, and then generated distributions of the hazard index. This was done to account for differences in pharmacokinetics and potency among different phthalates. This is also necessary to estimate upper bound risk accurately, that is, to avoid summing 95th percentile exposures from individual phthalates.
    b. Is this approach to cumulative risk assessment appropriate and scientifically defensible?
    c. Are there alternative approaches you would recommend that the CHAP consider?

3. **Critical Effect and Reference Doses**
    a. The risk assessment focuses on male developmental effects. Is the choice of male developmental effects as the critical effect supported by the available animal and human data?
    b. Is it appropriate to regard male developmental effects in rodents as the critical endpoint for the cumulative risk assessment of phthalates in humans?
    c. Is there sufficient understanding of the phthalates' mode of action to extrapolate from male rat developmental effects to humans?
    d. In case 1, is the selection of published reference doses for individual phthalates (Kortenkamp and Faust, 2010) appropriate for this task? (See section 2.7.2.2).
    e. In cases 2 and 3, is the derivation of the individual phthalate reference doses appropriate, including selections of studies, endpoints, and uncertainty factors? (See section 2.7.2.2).
    f. Are there other endpoints that should be considered for risk assessment (either for individual phthalate risk assessments or cumulate risk assessment)?

4. **Sensitive Populations**
    a. Is the selection of sensitive populations appropriate?
    b. Are "women of reproductive age" the most sensitive population?
    c. Does the risk assessment methodology adequately address the potential risks to children?
    d. In case 3, the CHAP derived RfDs specific for antiandrogenic effects in male offspring exposed perinatally. These RfDs are not necessarily the most sensitive endpoints for a given phthalate. Is it appropriate to apply reference doses based on prenatal exposure to infants or other populations?

5. **Scope. Did the CHAP adequately address their charge, as outlined above?**

6. **Please comment on the overall conclusions and recommendations.**

7. **Research Needs**
    a. Are the CHAP's recommendations for future research appropriate?
    b. Are there any other suggestions for future work that would reduce the uncertainty in the risk assessment?

8. **Are there any other scientific issues or comments on the report?**

Appendix B

Pre-Review Teleconference Slides and Additional Information for Reviewers

This page left intentionally blank.

Reviewer Clarifying Questions from Pre-Review Teleconference

Reviewers asked a number of clarifying questions during the teleconference. Either TERA or CPSC staff responded to each question.

1) One reviewer noted he had presented information to the CHAP during a public meeting and asked if this was a conflict of interest TERA responded that they did not consider that activity to be a conflict of interest with this review. CPSC noted that all the CHAP meetings and reports are available online.
2) A reviewer asked about the literature search. CPSC explained that the literature search strategy used is described in the report.
3) Were only urine biomonitoring data considered? CPSC responded that urinary and other data were considered.
4) Were different phthalates measured in each sample (concurrent exposure)? CPSC said, yes, e.g., NHANEs data.
5) Are reviewers expected to review all sections of the report? TERA responded that each reviewer should focus on the sections corresponding to their expertise.
6) How is level of "no harm" defined? CPSC noted that Congress was referring to that used in FQPA and the RfD is meant to be a negligible risk level.
7) Did the CHAP authors cite any data that were not peer reviewed? CPSC explained that some data were not peer reviewed (e.g., TXCA 8e and gray literature), but that for the most part they used peer-reviewed literature and gave greater weight to peer-reviewed literature.
8) A reviewer asked about the quality and general representativeness of the exposure data, and whether non peer-reviewed literature was included. CPSC noted that these data are in Appendix E and that the exposure data are uneven - some are good and other data are lacking.
9) Did the CHAP consider the issue of compounding conservatism from potential use of high levels of multiple phthalates? CPSC noted that for each scenario the average and the 95^{th} and 99^{th} percentiles were used for each exposure route.
10) How was child defined? CPSC explained that child was defined as prenatal to adulthood. For the exposure assessment, the age categories are defined.

Additional Questions and Information Provided to Peer Reviewers during the Course of the Review

During the course of the peer review the experts were invited to ask questions or ask for additional information that would aid them in their review. Two clarifying questions were received from individual reviewers. Below are the questions and the responses provided by CPSC staff.

1) *Reviewer Question:*
   I can't figure out Table 2.16 on p 56 of the main document. It is one of the most

important tables. I can assume that the row that says HI is the HI and all rows above are the HQ. However, take a look at DIBP row. Under NHANES, there is "1" "2" and "3". Under 1, there are 3 values, presumably HQ of 0.001, 0.01, and 0.01. The table title says that the median, 95th and 99th percentiles are provided. So, which is the median? There are 9 numbers under NHANES and DIBP. Also, the bottom row says % with HI>1.0. Percentages of what? It would help if they could identify what the 1, 2, and 3 mean and identify why each "box" (i.e., DIBP, NHANES, 1) has 3 numbers.

Response from CPSC Staff:
1, 2, & 3 are cases 1, 2, & 3. Each case is a different set of RfDs.

The hazard quotients are the median (on top), 95th percentile (middle), and 99th percentile (bottom). The same applies to the hazard index.

For example, DIBP, NHANES pregnant women, case 1. The median HQ is 0.001, 95th percentile is 0.01, and the 99th percentile is 0.01.

The % with HI>1 is the percentage of the population with a hazard index greater than 1. The population is either NHANES pregnant women, SFF pregnant women (prenatal); SFF women postnatal; or SFF infants.

2) *Reviewer Question:*
I wonder who the audience is [for the CHAP report]. Obviously, the Executive Summary is for exec's, congressional staffers, etc. I also expect that the appendices are more for the scientists. However, what about the main text? I ask because some places hold the reader's hand, but other places assume a fair amount of knowledge about the science and process of risk assessment.

Response from CPSC Staff:
The official audience is "the Commission," including the staff and the five appointed officials who will vote on any subsequent regulations.

Generally, CHAP reports, contractor reports, and CPSC staff reports, are written for a scientific or technical audience, keeping in mind that the readers may be from a range of disciplines. The CPSC staff will brief the Commissioners on the report and will provide any explanation as needed.

In some sections of the report, there are explanations for the benefit of scientists who might not be familiar with risk assessment, or who might not be experts in development.

I attached copies of two previous CHAP reports [DEHP and DINP], if you think that might be helpful.

Pre-Review Teleconference Slides

Chronic Hazard Advisory Panel on Phthalates and Phthalate Substitutes— Peer Review of the Draft Report

U.S. Consumer Product Safety Commission
June 4, 2013

THIS INFORMATION IS DISTRIBUTED SOLELY FOR THE PURPOSE OF PRE-DISSEMINATION PEER REVIEW UNDER APPLICABLE INFORMATION QUALITY GUIDELINES. IT HAS NOT BEEN FORMALLY DISSEMINATED BY THE CONSUMER PRODUCT SAFETY COMMISSION. IT DOES NOT REPRESENT AND SHOULD NOT BE CONSTRUED TO REPRESENT ANY AGENCY DETERMINATION OR POLICY.

DRAFT--FOR OFFICIAL USE ONLY 1

Consumer Product Safety Improvement Act (2008)

- Effective February 2009
- Permanent Ban (>0.1%)
  - Dibutyl phthalate (DBP), butyl benzyl phthalate (BBP), & di(2-ethylhexyl) phthalate (DEHP)
  - Children's toys and child care articles
- Interim Ban (>0.1%)—Pending Review by CHAP
  - Diisononyl phthalate (DINP), diisodecyl phthalate (DIDP), & di-n-octyl phthalate (DNOP)
  - Children's toys that can be placed in a child's mouth and child care articles
- CHAP on Phthalates & Substitutes
- Third Party Testing, Certification (§102)

DRAFT--FOR OFFICIAL USE ONLY 2

Chronic Hazard Advisory Panel (CHAP)

- Consumer Product Safety Act (CPSA)
- Cancer, birth defects, & gene mutations
- Seven independent scientists:
  - Selected by CPSC from a list of 21 nominated by National Academy of Sciences
  - Possess the required expertise
  - Not employed by the federal government, except NIH, NTP, or NCTR
  - Not associated with manufacturers
- Members select a Chair and Vice-Chair

CHAP's Charge (CPSIA)

For all phthalates used in children's products consider:

- All potential effects on children's health, including endocrine disruption
- Individual and cumulative risks
- Estimated exposure to children, pregnant women, and others
- Total phthalate exposure from:
  - Children's products
  - Personal care products
  - All other sources
- All routes of exposure

CHAP's Charge (continued)

- Using appropriate safety factors, derive a level of no harm to:
  - Children
  - Pregnant women
  - Other susceptible individuals
  - Offspring
- Phthalate alternatives used in children's products
- Conducted *de novo*, using
  - All available information
  - Objective methods
- Recommend to CPSC whether to ban any additional phthalates or phthalate alternatives

DRAFT--FOR OFFICIAL USE ONLY 5

CHAP Meetings

- First meeting April 2010
- Seven meetings in Bethesda, MD (all public)
- Six teleconferences (all public)
- Public testimony (July 2010)
- Ten invited experts presented to the CHAP:

- Kim Boekelheide, Brown
- Tom Burke, Johns Hopkins
- Paul Foster, NIEHS
- Earl Gray, EPA
- Matt Lorber, EPA
- Jeff Peters, Penn State
- Richard Sharpe, MRC
- Richard Stahlhut, U. Rochester
- Jamie Strong, EPA
- Shanna Swan, U. Rochester

Additional information on CHAP meetings and data submitted to the CHAP is available at Chronic-Hazard-Advisory-Panel-CHAP-on-Phthalates/

DRAFT--FOR OFFICIAL USE ONLY 6

Phthalates in the CHAP Report

Fourteen (14) Phthalates in the Report

- CPSIA phthalates (6)
  - DBP, BBP, DEHP, DINP, DNOP, DIDP
- Phthalates from biomonitoring studies (2)
  - Dimethyl phthalate (DMP), diethyl phthalate (DEP)
- Phthalates affecting male development (5)
  - Diisobutyl phthalate (DIBP), di-*n*-pentyl phthalate (DPENP), di-*n*-hexyl phthalate (DHEXP), dicyclohexyl phthalate (DCHP), diisooctyl phthalate (DIOP)
  - Straight-chain phthalates with 3 to 6 carbons and certain branched or cyclic phthalates affect male development
- Increasing exposure—di(2-propylheptyl) phthalate (DPHP) (1)

Phthalate Alternatives in the CHAP Report

- Six (6) Phthalate Alternatives
  - 2,2,4-Trimethyl-1,3 pentanediol diisobutyrate (TPIB)*
  - Di(2-ethylhexyl) adipate (DEHA)
  - Di(2-ethylhexyl) terephthalate (DEHT or DOTP)*
  - Acetyl tributyl citrate (ATBC)*
  - 1,2-Cyclohexanedicarboxylic acid, diisononyl ester (DINX)*
  - Tris(2-ethylhexyl) trimellitate (TOTM)
- Currently used or likely to be used in children's products

* Known to be used in children's products

Health Effects of Phthalates

After considering all health effects of phthalates, the CHAP decided to:

- Focus on male developmental effects—phthalate syndrome in rats (NRC 2008)
- Reduced AGD, retained areola/nipples, cryptorchidism, hypospadias…
- These effects are due, in large part, to inhibition of testosterone synthesis—antiandrogenicity
- Fetus > juvenile > adult
- Mixtures of antiandrogenic phthalates and other antiandrogens are dose additive (Howdeshell et al. 2008)

Epidemiology

- Phthalate syndrome resembles testicular dysgenesis syndrome (TDS) in humans
  - TDS includes cryptorchidism, hypospadias, poor sperm quality, & testicular cancer
- Phthalate exposure associated with:
  - Reduced AGD in neonates
  - Cognitive & behavioral effects in children
  - Reproductive effects in adult males
- Epidemiology studies limited by study design & concomitant exposure to multiple phthalates

Human Health Risk Assessment

- Based on animal studies, the most sensitive target is the fetus, followed by neonates
- Thus, the CHAP considered risks to:
  - Women of reproductive age
  - Pregnant women
  - Infants
- Dose-response assessment based on animal data

Exposure Assessment

The CHAP assessed exposure by two approaches:

- Human biomonitoring—total exposure
  - NHANES—women of reproductive age & pregnant women (n ≈ 1200)
  - Study for Future Families (SFF)—mothers (pre- & postnatal) and infants (n ≈ 300)
- Modeling—exposure by source
  - Diet
  - Consumer products
  - Cosmetics
  - Environment

Cumulative Risk Assessment

- Hazard Index Approach
- Biomonitoring data for five antiandrogenic phthalates—DEHP, DBP, DIBP, BBP, & DINP
- Three sets of RfDs (phthalate syndrome endpoints)
  - Cases 1 & 2 published values
  - Case 3 derived by CHAP
- Calculated hazard quotients & hazard index for each individual in a population
- Generated distribution of HI values
  - Avoids summing 95th percentile exposures

Key CHAP Findings—Hazard Index

- Women of reproductive age/pregnant women—up to 10 % have HI > 1
- Infants—about 5 % have HI > 1
- HI primarily due to DEHP:
  - Cases 1 & 3 > 90% from DEHP
  - Case 2 > 50% from DEHP
- Median hazard quotients generally ≤ 0.02, except DEHP

Key CHAP Findings—Exposure Modeling

Women of Reproductive Age

- Most exposure from diet
- Cosmetics/personal care products (DEP & DBP)

Infants & Toddlers

- Most exposure from diet
- Cosmetics/personal care products (DEP)
- Mouthing teethers & toys (if phthalates allowed)

DRAFT--FOR OFFICIAL USE ONLY 15

CHAP's Recommendations—Criteria

- Antiandrogenic Phthalates
  - Cumulative risk > risk in isolation
- Non-antiandrogenic Phthalates & Phthalate Alternatives
  - Most sensitive health endpoints
  - Risk in isolation

DRAFT--FOR OFFICIAL USE ONLY 16

CHAP's Draft Recommendations

- Permanent ban—DIBP, DPENP, DCHP, DHEXP, DINP
  - These are antiandrogenic
  - DBP, BBP, & DEHP are already permanently banned
- Interim ban—DIOP (limited data)
- Remove from interim ban—DNOP, DIDP
- No action
  - Phthalates—DMP, DEP, DPHP
  - Alternatives—ATBC, DEHA, DEHT, DINX, TOTM, TPIB
- Specific data needs
  - Phthalates—DIOP, DPHP
  - Alternatives—ATBC, DEHA, DEHT, DINX, TOTM, TPIB