Health Sciences

Exponent®

Review of Phthalate Esters Epidemiology

Review of Phthalate Esters Epidemiology

Prepared for:
Chronic Advisory Health Panel (CHAP)
Of the Consumer Products Safety Commission
At the request of the
Phthalate Esters Panel of the American
Chemistry Council

Prepared by:
Jane Teta, Dr.PH, MPH
Jeanne Manson, Ph.D.
Meghan Wagner, MPH
Exponent
420 Lexington Avenue, Suite 1740
New York, New York 10170

July 2010

1003844.000 A0T0 0710 JT01

Over the past decade, an increasing number of epidemiology studies have appeared in the scientific literature investigating *in utero* biomarkers of phthalate exposures and subsequent birth outcomes and developmental effects in infants. Several studies investigated associations with gestational age, while one study each considered an association with cryptorchidism (or undescended testicles) and anogenital distance (AGD). A number of studies have focused on neurodevelopmental effects, such as reduced cognitive function and adverse behavioral outcomes, using biomarker phthalate metabolite exposure measurements at older ages. A few other studies have explored possible hormonal disruption using phthalate metabolite measurements in maternal breast milk or from children's urine samples. In general, these studies have a number of features in common: a proposed hypothesis based on the results of animal studies, enumeration of study limitations, a discussion of possible mechanisms for any associations observed, and, routinely, a call for more research.

Virtually all of the studies report one or more statistically significant associations. Causal interpretation of these findings is problematic, however, due to several limiting factors, often well described by the investigators. I will discuss the strength of the epidemiologic evidence in light of the challenges that researchers have faced that include: exposure misclassification, inter-study inconsistencies, lack of replication, residual confounding, weak associations, and multiple comparisons. A more detailed description of each study (including its strengths and weaknesses) has been provided in our full report.

1. **Exposure Misclassification**: Most authors raise the issue of whether their study captured the appropriate window of exposure, given the short half lives (less than 24 hours) of phthalate biomarkers (Koch et al., 2005). DEHP metabolites were shown not to be highly reproducible during the last six weeks of the third trimester by Adibi et al. (2009). Other studies, however, suggested relative stability of phthalate biomarkers for weeks or months among men and non-pregnant women (Hoppin et al., 2002; Hauser et al., 2004; Teitelbaum et al., 2008). The metabolism of pregnant women changes rapidly and they have a 30% increase in circulating blood volume. Of particular concern are cross-sectional neurobehavioral studies that attempted to link spot urines of elementary school children to parental or teacher neurobehavioral performance ratings. Wolff et al. (2008, 2010) appropriately call for a more comprehensive, integrated exposure assessment prenatally and before puberty.

2. **Inter-study inconsistencies**: Where there are multiple studies of the same endpoint, the results have not demonstrated reasonable consistency. This is particularly the case for studies that have examined gestational age at delivery (Lantini et al., 2003; Wolff et al., 2008; Adibi et al., 2009; Meeker et al., 2009; Whyatt et al., 2009). Whyatt et al. (2009) reported a statistically significant inverse relationship of DEHP metabolites and gestational age at delivery. This finding is consistent with that of Lantini et al. (2003) who measured MEHP in cord blood. It is inconsistent, however, with Wolff et al. (2008) who reported statistically significant positive associations of "low" molecular weight phthalate (MWP) metabolites[1] and MEHP with gestational age. Adibi et al. (2009) also

[1] Low molecular weight phthalates ("low" MWP) and high molecular weight phthalates ("high" MWP) are defined according to the definitions of the researchers whose study is being discussed.

reported increasing gestational age with increasing DEHP metabolites. It is noteworthy that, although their results were divergent, both the Whyatt et al. (2009) and Wolff et al. (2008) studies were conducted among African-American and Hispanic New York City mothers, with 311 mothers in the former study and 404 in the latter. Metabolite concentrations were also similar in these two studies. Meeker et al. (2009) reported a positive odds ratio of borderline statistical significance for pre-term birth for the metabolite MEHP, comparing cases and controls with greater than median levels, while Adibi et al. (2009) reported an inverse association for pre-term birth and DEHP metabolites. In a nested case-control study of cryptorchidism, Main et al. (2006) reported no differences in any metabolites measured in mothers' milk or differences in gestational age between boys with undescended testicles and controls.

The two neurobehavioral studies by Engel et al. (2009, 2010), one in newborns and one when these children reached 4-9 years of age, both reported sex-specific effects. Newborn girls showed a significant linear decline in Orientation and Quality of Alertness, with increasing concentrations of "high" MWP. Boys showed a different pattern with some indication of better motor performance with higher concentrations of "low" MWP. In the study of older children, however, boys exhibited poorer scores with increasing concentrations of "low" MWP.

3. **Lack of replication**: First time reports of associations, particularly studies employing novel methodology, require replication before a causal association can be considered seriously. For example, there have been no studies of anogenital distance (AGD) and exposure to phthalates in humans, other than the one by Swan et al. and her colleagues (2005) and a subsequent study by Swan alone (2008). Studies that examined similar outcomes are not true replications if the source of exposure is substantially different, such as studies where exposures were sampled from urine versus cord blood or breast milk.

4. **Residual confounding**: Errors in measurement or inadequate surrogates for confounders can result in residual confounding and bias risk estimates in either direction, that is, an under- or over- estimation of risk. The existing studies have collected information on numerous variables that may be confounders, mostly from parent questionnaires. There is a long list of covariates that are known or suspect risk factors for developmental effects in children, for example, age, sex, ethnicity, race, and mothers' pre-pregnancy height, weight, body mass index (BMI), smoking history, education, IQ, marital status, and history of asthma, hypertension, and diabetes. To confound an association, a risk factor also would have to be associated with exposure to phthalates. Main et al. (2006) note the possibility of unknown factors related to phthalate exposure. While knowledge is incomplete on the distribution of phthalate exposures, a few important characteristics have emerged. Minorities, those of lower socio-economic status, and obese persons typically have higher levels of phthalate metabolites in their bodies. Many of the study authors tested for confounding with these variables by examining the effect with and without the factor in the statistical model. The effectiveness of this approach relies on the accuracy of parental recall. For example, Wolff et al. (2008) reported a positive but small statistically significant correlation (r=0.18) between "low" MWP metabolites and BMI, but raise the issue of crude estimates of maternal anthropometric features based on

maternal self-reports.

It has been estimated that 90% of DEHP intake (except for intake in infants) is from food (Kavlock, 2006). The most influential food products are fats, oils, and dairy products. If these food products are independent predictors of risk for developmental effects or hormonal alterations, they could also be confounders in an association between DEHP and developmental effects. While most studies that have examined this association control for BMI or body weight, there is uncertainty as to whether these variables adequately control for potential dietary confounding. No study has attempted to control specifically for dietary intake.

5. **Weak associations**: Weak associations are more readily explained by confounding or other forms of bias than stronger associations. Many of the associations observed in these studies were not statistically significant and are, therefore, assuming no error, chance findings or insensitive due to small sample sizes. Where statistically significant associations have been observed, they are generally of low magnitude, which could easily be due to study limitations.

   Another challenge to interpretation with weak associations is whether they are clinically relevant. The endpoints examined in the developmental studies are of unknown significance, either because of the subtle changes observed or because the relationship to a clinical diagnosis is absent. Seven years ago, Latini et al. discussed the need for a clearer understanding of the clinical relevance of their measured associations between phthalate metabolites and shorter gestational age. In 2008, Wolff et al. noted the small effect sizes related to longer length of gestation. Adibi et al. (2009), who also reported longer length of gestation associated with higher levels of phthalate metabolites, commented "The clinical or population significance of 2-3 days in gestational length is difficult to evaluate." In the Wolff et al. (2010) study, the weak associations between "low" MWP biomarkers with the timing of puberty (as measured by breast development and pubic hair) in girls was in contrast to the inverse relationship with "high" MWP biomarkers. None of the adjusted prevalence ratios was statistically significant and there were no notable differences from the null, i.e., 1.0. Engel et al. (2010) reported poorer parent-rated behavioral and executive functioning in boys, not girls, associated with "low" MWP, not "high" MWP, although few children "met the standard at risk or clinically significant criteria." Main et al. (2006), on the other hand, investigated a well-defined clinical entity, cryptorchidism, and reported no association with any of the six phthalate metabolites measured in mothers' milk.

6. **Multiple comparisons**: Since the studies are exploratory in nature, many of them have analyzed multiple phthalate metabolites and multiple outcomes. Given a large number of comparisons, some associations will be statistically significant just due to chance. Meeker et al. (2009) referred to the "large number of statistical comparisons." Wolff et al. (2008, 2010) noted that their statistically significant associations may be due to multiple comparisons. Multiple comparisons are particularly problematic for studies of neurodevelopmental effects. Engel et al. (2010), for example, included behavioral rating inventories of 86- and 130-item questionnaires that were examined separately for boys

and girls and “high” and “low” MWP, resulting in over 800 comparisons. Swan et al. (2010) included over 500 comparisons while examining reduced masculine play in boys.

Biomarker studies of developmental effects pose numerous challenges. The criteria for causation, such as consistency, biologic plausibility, magnitude of risk, and even temporality (as several studies are cross-sectional design), are far from satisfied across the studies of phthalate metabolite exposures and developmental effects. Those few studies suggesting a dose-response are uncertain due to questions of residual confounding and exposure misclassification. Because of these challenges, authors are appropriately cautious in their interpretation of the observed weak and inconsistent associations.