Supplemental Material

Sexual Dimorphism. Participants concurred that when selecting sex hormone-dependent endpoints, the sexual dimorphism of candidate outcomes should be considered. The more distinct the dimorphism, the easier it is to observe perturbations caused by endocrine-active exposures. Parameters selected should be those that can be measured reliably, are logistically feasible and for which there is empiric evidence of hormone-responsiveness. Recommended measures include thigh fat (greater degree of sexual dimorphism than other areas; skinfold better than circumference (Antoszewska 1992)), and breast size (regresses after birth, therefore a hormonally responsive tissue). According to research from Swan et al. (2005), anogenital distance (AGD) is also recommended. Hand digit length is difficult to accurately and reproducibly measure since babies do not like to lay their hands flat.

Nutritional Assessments

Research studies assessing endocrine-sensitive endpoints within the first years of life are encouraged to include a nutritional assessment component. The purpose of the nutritional component is three-fold: (1) to assess dietary intake and/or nutritional status of specific nutrients that play a role in the endocrine system such as iodine, vitamin D, and lipids; (2) to identify and/or attribute sources of exposure to endocrine disruptors which may also be a source of nutrients; and (3) to document potential changes in food intake that might be relevant to endocrine-sensitive endpoints. For example, nausea and

vomiting during pregnancy may result in transient changes in maternal diet. Ideally, prenatal maternal assessments during the 1st and 3rd trimesters and during early lactation would be made as well infants' assessment during early lactation, at 6 and 12 months. In addition, the father's diet during the three months prior to conception may also be an important co-factor.

The choice of dietary assessment method depends on the study objectives, and the time and resources available for the study (Fehily and Johns 2004). Workshop participants recommended the Food Frequency Questionnaire. However, there are few dietary assessments targeted and validated for infants and toddlers (Parrish et al. 2003; Rockett and Golditz 1997). There is also uncertainty about whether parents can accurately recall their child's dietary behaviors, particularly since children are often fed by other caregivers (Parrish et al. 2003). For measuring endocrine-sensitive endpoints, it is important to collect the duration and exclusivity of breast feeding, the brand of infant formula, any changes in type of formula, reason for the change, and also the reason why the infant is on formula.

Biomarkers of endocrine-related nutrients may be employed to ascertain nutrient intakes or measure nutrient status. Their use is often limited by high collection and analysis costs as well as concerns related to invasiveness, especially in infants. These drawbacks suggest that large studies with resource constraints should employ biochemical indicators to validate questionnaires. For example, a serum radioimmunoassay for vitamin D may be compared with vitamin D status based on dietary intake (Weiler et al. 2007).

Erythrocyte fatty acids (especially long chain polyunsaturated fatty acids) which serve as a biochemical indicator of long term intake may be useful to assess status and validate questionnaires (Cao et al. 2006; Sullivan et al. 2006; Weiler et al. 2005). Although iodine can be measured in a casual sample of urine from the newborn, iodine is known to fluctuate among individuals in the short term but is a good indicator of iodine sufficiency in the population (Hambidge 2003). Maternal and infant finger nail clippings may be used to measure body burdens of other relevant nutrients such as selenium; however, the nails must be carefully washed, the growth rate is variable, and multiple clippings may be necessary to have sufficient mass for analysis (Longnecker et al. 1993).

Research Needs

More data are needed on whether there are nutrient differences, particularly among the water-soluble phytoestrogens measured in maternal blood at delivery, which vary according to length of labor and delivery. Thus, there is a need to develop and validate questionnaires to estimate intake of phytoestrogen and hormones along with the food sources of exposure.

Neurodevelopment

Although neurodevelopment was recognized as an important endocrine-sensitive endpoint, this assessment was not a major focus of the workshop and consequently the following summary of discussions is not meant to represent the breadth of research and

assessment tools in this area. It should also be noted that the appropriate neurological assessment to use varies by age, culture and language.

The brain is developing both pre- and postnatally and different regions of the brain develop at different rates and times. These regions may be differentially susceptible to the influence of environmental contaminants. Moreover, the disabling effects of a teratogenic substance may not emerge at birth, but rather when the potentially vulnerable neural or behavioral system has matured or compensatory mechanisms are no longer adequate.

Sensory testing in infancy, including both visual and auditory assessment, was identified as a sensitive endpoint of developmental neurotoxicity testing. These measures are not only susceptible to neurotoxicity, but are also important neurological systems that feed into higher cognitive processes. For visual functioning, contrast sensitivity (i.e. a measure of contrast detection for a range of spatial frequencies) was considered to be a better measure than a visual acuity measurement, and deficits have been shown in infants exposed gestationally to organic solvents (Till et al. 2005). Color-vision testing has also proven to be clinically useful, and may detect neurological changes that precede changes in more standard measures of visual function, such as visual acuity. Electrophysiological measures, such as visual evoked potentials, can be employed to test vision in infants. .

Hearing is also very vulnerable to environmental toxicants. Testers must be very well trained in scoring and interpretation of the tests (especially for babies).

Global measures like IQ or a developmental index on the Bayley scale are often not sensitive to detecting subtle neurodevelopmental effects of a chemical, particularly at early stages of development. A neurodevelopmental assessment should target abilities that tap into areas in the brain where effects are hypothesized to occur; however, these may not be amenable to testing in very young children or may not come 'online' until a later age once the child has acquired certain competencies. Animal research can help direct which region of the brain to study and the appropriate endpoint to study in terms of cognitive ability.

For behavior, there are many high quality sensitive questionnaires, such as the Child Behavior Check List which can be administered in children as young as 18 months of age. The Connor's questionnaire is another widely-used measure that is relatively sensitive for assessing Attention Deficit Hyperactivity Disorder and Fetal Alcohol Syndrome.

For neurodevelopmental testing, it is important to measure potential confounders such as parental IQ and education, quality of the care-taking environment, and parental stress because they are known predictors of child neurodevelopment and may also confound the association between exposure and outcome.

One of the major challenges of deriving a sensitive battery of neurodevelopmental tests is determining which effects are most likely to be related to the exposures of interest. Often

investigators are dealing with multiple compounds that inherently have multiple modes of action. Thus, it is important to test a number of functions and cast a broad net to identify the potential range of deficits early in development. With multiple exposures, one chemical may impact one particular function, such as language, whereas another chemical may affect another function, such as memory. The drawback to the broad-domain approach is that there may be less specificity in the tests due to the limited amount of time available for the assessment. For toddlers with limited attention spans, the maximum length of the assessment is at best 2 hours. Early in development, basic behaviors, such as variability in performance and speed of performance, which contribute to the development of executive functions and memory in an older child can be assessed.

Behavioral testing in children 18 months of age is possible, but made difficult by the challenge of attracting and sustaining the child's attention over the length of the assessment. As mobility of the child increases, it is harder to engage their attention and training of testers becomes more critical. Since 3- to 6-month old infants can be easily engaged, the Bayley given as early as 3 months of age is ideal for measuring attention, while assessments at 24 and 36 months provide a more comprehensive overview of more advanced abilities.

Sex-related behaviors and toy choices may also be affected by hormones and chemicals and should be considered as endocrine-related sexual dimorphic endpoints. These behaviors can be observed in children as young as 24 months.

Spatial abilities and language acquisition show sexual dimorphism. The Mullen test is divided into scales and has measures of visual and spatial abilities. The Fagan test can be difficult to obtain and expensive, but relatively easy to administer, while the 3-month Bayley has many similar components to the Fagan test.

It is also important to consider gross measures of neurodevelopment while doing individual exams such as primitive reflexes of neonates (Young et al. 2005).

References

Antoszewska A, Wolanski N. 1992. Sexual dimorphism in newborns and adults. Stud Hum Ecol 10:23-38.

Cao J, Schwichtenberg KA, Hanson NQ, Tsai MY. 2006. Incorporation and clearance of omega-3 fatty acids in erythrocyte membrane and plasma phospholipids. Clinical chemistry 52:2265-2272.

Fehily AM, Johns AP. 2004. Designing questionnaires for nutrition research. Nutr Bull 19:50-56.

Hambidge M. Biomarkers of trace mineral intake and status. 2003. J Nutr 133:948S-955S.

Longnecker MP, Stampfer MJ, Morris JS, Spate J, Baskett C, Mason M, Willett WC. 1993. A 1-yr trial of the effect of high-selenium bread on selenium concentration in blood and toenails. Am J Clin Nutr 57: 408-413.

Parrish LA, Marshall JA, Krebs NF, Rewers M, Norris JM. 2003. Validation of a food frequency questionnaire in preschool children. Epidemiology 14:213-217.

Rockett HRH, Colditz GA. 1997. Assessing diets of children and adolescents. Am J Clin Nutr 65 (suppl 1):1116S-1122S.

Sullivan BL, Williams PG, Meyer BJ. 2006. Biomarker validation of a long-chain omega-3 polyunsaturated fatty acid food frequency questionnaire. Lipids 41:845-850.

Swan SH, Main KM, Liu F, Stewart SL, Kruse RL, Calafat AM, Mao CS, Redmon JB, Ternand CL, Sullivan S, Teague JL, Study for Future Families Research Team. 2005. Decrease in anogenital distance among male infants with prenatal phthalate exposure. Environ Health Perspect 113:1056-1061.

Till C, Westall CA, Koren G, Nulman I, Rovet JF. 2005. Vision abnormalities in young children exposed prenatally to organic solvents. Neurotoxicology. 26:599-613.

Weiler H, Fitzpatrick-Wong S, Schellenberg J, McCloy U, Veitch R, Kovacs H, Kohut J, Yuen CK. 2005. Maternal and cord blood long-chain polyunsaturated fatty acids are predictive of bone mass at birth in healthy term-born infants. Pediatr Res 58:1254-1258.

Weiler HA, Leslie WD, Krahn J, Steinman PW, Metge CJ. 2007. Canadian aboriginal women have a higher prevalence of vitamin D deficiency than non-aboriginal women despite similar vitamin D intakes. J Nutr 137:461-465.

Young JG, Eskenazi B, Gladstone EA, Bradman A, Pedersen L, Johnson C, Barr DB, Furlong CE, Holland NT. 2005. Association between in utero organophosphate pesticide exposure and abnormal reflexes in neonates. Neurotoxicology 26:199-209.