

## Dr. Abby Li's Response to questions on the draft Toxicological Review

Overall, the presentation of data in tables including absence and presence of findings, dose levels, sample size was very helpful as a tool/guide for reviewing the available scientific literature and understanding EPA's rationale for selection of critical endpoints for risk assessment purposes. The downside of these tables is that it's more difficult to get a sense of the consistency across findings within a study, but the supplemental information provided additional perspective. The EPA team should be commended for the major effort that it takes to synthesize a vast amount of data points into useful summary tables.

The comments in this response are based primarily on evaluation of key developmental neurotoxicity animal studies for risk assessment purposes. A primary concern is that there are important criteria that may not have been fully considered in assessing the quality and utility of studies for risk assessment purposes.

### **1. Literature search/study selection and Evaluation.**

The process for identifying and selecting pertinent studies for consideration in developing the assessment is detailed in the *Literature Search Strategy/Study Selection and Evaluation* section.

#### **a. Please comment on whether the literature search approach, screening, evaluation, and selection of studies for inclusion in the assessment are clearly described and supported.**

The EPA did a thorough job documenting search terms used to identify studies in the main and supplementary report. The first 2 dotted line boxes of excluded references in Figure LS-1 were self-explanatory. However, the criteria used to exclude the 600 references in the manual screen of manuscripts (3<sup>rd</sup> dotted line box) are less clear-cut. It is appropriate to exclude papers that are "not relevant to B(a)P toxicity in mammals", or have "inadequate reporting of study methods or results" or "inadequate basis to infer exposure". However, it's not clear what EPA's definitions for "relevant" or "inadequate" are. EPA could be using Section 3.2 and 3.3 and elements of Section 4.2 and Section 6 of the Preamble as the basis for what's adequate reporting of study methods. If so, EPA may want to reference these sections from the Preamble. If studies were excluded due to inadequate reporting of study methods or results, it may be appropriate to list the references in the supplementary information for greater transparency.

#### **b. Please comment on whether EPA has clearly identified the criteria (e.g. study quality, risk of bias) used for selection of studies to review and for the selection of key studies to include in the assessment.**

The preamble (section 4.2, p xx) refers to EPA guidelines for further guidance on the nuances of evaluating experimental studies for developmental toxicity, reproductive toxicity and neurotoxicity. Important criteria from these guidelines relevant to

evaluating key endpoints selected for B(a)P include (a) blind observations; (b) counterbalancing the time of testing across dose levels; (c) operational definitions for subjective measures; (d) sample size for behavior is 10 males and 10 females from 20 litters (1 pup/litter); (e) the litter is the required experimental unit of analysis . Section 4.2 of the EPA IRIS Preamble also mentions consideration of historical control and maternal toxicity in assessing the findings.

The B(a)P assessment did not consistently evaluate the studies for these characteristics. As discussed in greater detail below, EPA's criteria for evaluating studies for risk assessment purposes may not have taken into consideration important methodological issues that can impact the dose-response assessment.

**c. identify any additional peer-reviewed studies from the primary literature that should be considered in the assessment of noncancer and cancer health effects of benzo[a]pyrene**

None have been identified at this point.

**2. Hazard identification. In section 1, the draft assessment evaluates the available human, animal, and mechanistic studies to identify the types of toxicity that can be credibly associated with benzo[a]pyrene exposure. The draft assessment uses EPA's guidance documents (see <http://www.epa.gov/iris/backgrd.html/>) to reach the following conclusions.**

**2a. Developmental toxicity (sections 1.1.1, 1.2.1). The draft assessment concludes that developmental toxicity and developmental neurotoxicity are human hazards of benzo[a]pyrene exposure. Do the available human, animal and mechanistic studies support this conclusion?**

An assessment of whether B(a)P is a "known" human developmental neurotoxicity hazard at exposure levels relevant to the general population (e.g. not worker exposed population) requires more critical evaluation of the epidemiology data, which is not the focus of this current response.

The animal literature suggests that developmental toxicity and developmental neurotoxicity are potential hazards of B(a)P exposure at oral doses of 0.02 mg/kg/day and higher. However, there are important experimental design weaknesses that suggest that these oral developmental neurotoxicity studies need to be repeated before these data can be considered for risk assessment purposes.

The most sensitive endpoint is based on Chen et al. (2012). This is a good quality study from the perspective of executing behavioral endpoints (e.g. blind observations, randomizing order of testing litters), but is severely confounded by the rotation of dams every 2-3 days "to distribute any maternal caretaking differences randomly across litters **and treatment groups**" (p.249 of original paper). This indicates that the dams were rotated to pups exposed to all 4 dose levels. It is not until the discussion section that we discover that the study used a "**within-litter**

**design**” explaining how the dams were rotated “across treatment groups”. Chen et al. (2012) correctly acknowledge that “this study design increases the risk of cross-contamination among groups, and untreated controls may also dominate the litter, and/or treated rats may be weak and subsequently rejected by the dams.” This important weakness was not identified by EPA, yet it severely diminishes the use of this study for dose-response risk assessment purposes. The frequent dam rotation could also introduce stress on the pups especially during the early weaning period. The assumption that maternal caretaking differences are randomly distributed across litters is unverified. It presumes that pups are equally vulnerable across the entire period of weaning. In fact, a dam exhibiting poor maternal care will have greatest influence on litters during the first 2 rotations of dams, and will be more likely to favor control pups over the treated rats.

Bouayed et al. (2009a) studied the effects of lactational exposures to offspring on a large number of behavioral endpoints. The exposed dams were also evaluated for maternal behaviors. It is a weaker study compared to Chen et al. (2012) because 5 litters/dose group is not adequate for behaviors measured. There is no mention of whether the observers were blind to treatment, or if the time of testing was balanced across dose groups. The major weakness of this study is that there was oversampling by testing 4 pups/litter, and the authors analyzed the data with n=20 pups without including litter as a factor in the statistical analyses. Although a good pilot study, this study is inadequate for risk assessment purposes. Care is also needed in interpreting the results of the elevated plus maze in relation to anxiety in humans. The elevated plus maze has been used as an initial screening tool anti-anxiety-like activity of chemicals, but equating increases in time in the open field directly with decreased anxiety is a hypothesis requiring further testing.

**2e. Other types of toxicity (section 1.1.4). The draft assessment concludes that the evidence does not support other types of noncancer toxicity as a potential human hazard. Are there other types of noncancer toxicity that can be credibly associated with benzo[a]pyrene exposure?**

No, based on EPA’s review of the literature.

3. **Dose-response analysis.** In section 2, the draft assessment uses the available human, animal, and mechanistic studies to derive candidate toxicity values for each hazard that is credibly associated with benzo[a]pyrene exposure in section 1, then proposes an overall toxicity value for each route of exposure. The draft assessment uses EPA’s guidance documents (see <http://www.epa.gov/iris/backgrd.html/>) in the following analyses.

**3a. Oral reference dose for effects other than cancer (section 2.1). The draft assessment proposes an overall reference dose of  $3 \times 10^{-4}$  mg/kg-d based on developmental toxicity during a critical window of development. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis, calculating points of departure, and applying uncertainty factors? Does the discussion of exposure scenarios (section 2.1.5) reflect the scientific considerations that are inherent for exposures during a critical window of development?**

As discussed above in great detail, the selection of the Chen paper for dose response assessment is not recommended given the within-litter design and possibility for cross-contamination.

4. **Executive summary. Does the executive summary clearly and appropriately present the major conclusions of the assessment?**
5. **Charge question on the public comments**

**In August 2013, EPA asked for public comments on an earlier draft of this assessment. Appendix G summarizes the public comments and this assessment's responses to them. Please comment on EPA's responses to the scientific issues raised in the public comments. Please consider in your review whether there are scientific issues that were raised by the public as described in Appendix G that may not have been adequately addressed**