

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA’s Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

**Revised Comments from Members of the Chemical Assessment Advisory Committee
Augmented for the Review of the EPA’s Draft IRIS Benzo[a]pyrene Assessment
(September, 2014)**

Comments Received as of April 24, 2015

Table of Contents

Dr. Scott Bartell.....	2
Dr. Ronald Baynes.....	6
Dr. Annette Bunge.....	10
Dr. Scott Burchiel.....	28
Dr. Anna Choi.....	30
Dr. John DiGiovanni.....	34
Dr. Joanne English.....	39
Dr. Michael Foster.....	47
Dr. Chris Gennings.....	51
Dr. Helen Goeden.....	54
Dr. Sean Hays.....	62
Dr. John Kissel.....	64
Dr. Ed Levin.....	66
Dr. Abby Li.....	70
Dr. Maureen Lichtveld.....	77
Dr. Barry McIntyre.....	78
Dr. Bhagavatula Moorthy.....	84
Dr. Miriam Poirier.....	88
Dr. Kenneth Portier.....	96
Dr. Steven Roberts.....	103
Dr. Richard Schlesinger.....	106
Dr. Leslie Stayner.....	108
Dr. Alan Stern.....	111
Dr. Charles Vorhees.....	114
Dr. Christi Walter.....	127

Dr. Scott Bartell

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

The criteria are clearly identified, and I am not aware of any additional studies that should be considered.

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?*

Yes, but on their own the findings from the epidemiologic studies are only weakly informative because they involved unmeasured exposures to mixtures of known developmental toxicants. It is not clear that the adverse health outcomes in these epidemiologic studies were due to benzo[a]pyrene exposure per se. Some of the wording in the assessment is too strong in this regard, e.g., "susceptibility to benzo(a)pyrene toxicity is indicated by epidemiological studies" on p. 1-20.

3a. Oral reference dose for effects other than cancer (section 2.1). *The draft assessment proposes an overall reference dose of 3×10^{-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?*

This RfD value is mostly well supported, but further clarification is needed regarding the use of NOAELs/LOAELs for the IgM and IgA endpoints from De Jong et al., 1999. The assessment states that the data for these endpoints were "inconsistent and non-amenable to dose-response modeling," but it is not clear what aspects of the data were inconsistent and whether those inconsistencies also cast doubt on the use of the data for deriving NOAELs/LOAELs. In addition, is it unclear why the data were non amenable to dose-response modeling. Was that

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

because of the inconsistencies, poor lack of fit, lack of converge of the fitting algorithm, or some other reason?

3b. Inhalation reference concentration for effects other than cancer (section 2.2). *The draft assessment proposes an overall reference concentration of 2×10^{-6} mg/m³ based on decreased fetal survival 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 2.2.5) reflect the scientific considerations that are inherent for exposures during a critical window of development?*

The RfC is mostly well supported, but the reason for avoiding dose-response modeling is entirely unclear. "Not amenable to BMD modeling due to the pattern of variability in the data set" (p. 2-17, lines 15-17) is vague and insufficient justification.

3c. Oral slope factor for cancer (section 2.3). *The draft assessment proposes an oral slope factor of 1 per mg/kg-d based on alimentary tract tumors in mice. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?*

Although consistent with previous methods used by EPA, choosing the maximum slope factor from the available studies is not the most reliable metric for characterizing carcinogenic potency; a mean or median would better capture the totality of evidence. The oral slope factor is otherwise well supported.

3d. Inhalation unit risk for cancer (section 2.4). *The draft assessment proposes an inhalation unit risk of 0.6 per mg/m³ based on a combination of several types of benign and malignant tumors in hamsters. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?*

Yes.

3e. Dermal slope factor for cancer (section 2.5). *The draft assessment proposes a dermal slope factor of 0.006 per ug/day based on skin tumors in mice. 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 scaling from mice to humans? Does the method for cross-species scaling (section 2.5.4 and appendix E) reflect the appropriate scientific considerations?*

Yes, this value is well supported, reflecting appropriate scientific considerations.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

OTHER COMMENTS

p. 1-1, line 20, "Two?" should probably be "Two."

p. 1-1, lines 25-29 are unclear. What exactly is meant by "independent effect on birth weight was not observed" and how does that differ from the 8% reduction in birth weight with ETS exposure in utero? The assessment should make it clear that this is a WTC-related exposure, and that the 8% reduction in birth weight was per doubling of adducts.

Table 1-1, no mention of potential confounding by other WTC-related exposures?

p. 1-36, lines 19-20 apply to the epidemiologic evidence for all health outcomes--all of the epi studies appear to use benzo(a)pyrene as a surrogate for exposure to complex PAH mixtures.

p. 1-47, lines 4-5 give the false impression that other epidemiologic cohorts described in this review use higher quality study designs, but many of the other cited studies seem to use one biomarker measurement per person, e.g. BaP-DNA adducts with a half-life of about 3-4 months. In such cases a statistical exposure model based on good occupational records may actually result in more accurate exposure assignment than a single exposure biomarker.

p. 1-82, lines 15-17, "the exposure-response patterns seen with benzo(a)pyrene measures make it unlikely that these results represent confounding by other exposures." Existence of an exposure-response pattern does not at all constitute evidence against confounding. This statement suggests a grave misunderstanding of confounding in observational studies, which can induce a strong exposure-response pattern with any non-causal agent (when correlated with an unmeasured causal agent).

p. 2-6, check on whether correlation of rats from same litter was accounted for in the dose-response modeling

p. 2-6, lines 16-17, kudos to EPA for requesting raw data (here and in other parts of the assessment).

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

p. 2-13, line 18, change "a statistically significance" to "statistical significance"

Table 2-3, why "low" confidence in immunological toxicity RfD versus "medium" for other effects?

p. 2-32, why were Brune study and Neal and Rigdon study not used to produce slope factors in this assessment? It would be helpful to summarize the reasons here and refer to the appropriate part of the assessment discussing their exclusion.

Dr. Ronald Baynes

Question #1. Literature Search Strategy / Study Selection

This review appears to have implemented a comprehensive search strategy to identify studies that are relevant to our understanding of benzo[a]pyrene toxicology using various approaches that involved *in silico* and *in vitro* methods as well as human and animal *in vivo* exposures. There are however, several references listed below that could have been added to this review to support their dermal assessment. This review excluded studies describing the use of therapeutic products containing BaP. While all of these products are formulations/mixtures containing BaP, it would have been useful to tabulate or provide an estimate of human exposure to BaP in these products; providing this information will be consistent with the current data in this Toxicology Review describing the adverse effects in human occupational exposure to BaP mixtures although the relative contributions of BaP and of other PAHs cannot be established.

Question #3e. Dermal Slope Factor

2.5.1. Analysis of Carcinogenicity Data (Choice of studies)

Principal Study

The Sivak et al (1997) appears to be the appropriate principal study. Three dose levels (0.0001, 0.001, 0.01%) and controls were evaluated with BaP in cyclohexane/acetone (1:1) were administered twice a week for 104 weeks. A multi-stage models were used to fit the mouse data; a POD for mouse data was 0.06 µg/day and the slope factor was 1.7 per µg/day for mouse. The mouse POD was scaled to a human POD and dermal slope factor was 0.006 per µg/day. There may be some issues with this approach which are discussed below.

Supporting Studies.

This toxicology review documents at least 10 carcinogenicity bioassay studies from 1959 to 1997 where they demonstrated a dose-response relationship and the document adequately presented why they were not selected as the principal study. Mice of various strains were the dominant species evaluated and various vehicles (e.g., acetone, toluene) were used as topical delivery of BaP. Mice appear to be the most sensitive species and should not be interpreted as a limitation of this dose response assessment. The slope factors ranged from 0.25 to 1.8 per µg/day compared to the 1.7 per µg/day for the principal study. These studies were not chosen for several good reasons such as limited dose response points, higher exposure levels, significant low-dose extrapolation, and incomplete exposure information.

2.5.4. Dermal Slope Factor Cross-Species Scaling

According to EPA guidance, if the fraction of the agent is absorbed from the diet for humans and animals differs, the U.S. EPA applies a correction when extrapolating the animal-derived value to humans. This needs to be taken into consideration when assessing dermal absorption especially when extrapolating from mice skin to human skin uptake.

Modeling of mouse tumor data generated a POD of 0.06 µg/day and a POD_{HED} of 17 µg/day. An interspecies scaling approach used $\frac{3}{4}$ power of body weight that was claimed to account for the more rapid distribution, metabolism and clearance in mice compared to humans. For many drugs and chemicals, absorption and penetration is significantly greater in mice skin than human skin. This approach assumes that mouse skin disposition can be scaled to an equivalent human skin disposition. The literature has sufficient human and mouse skin BaP absorption data (K_p, flux, etc) and metabolism data that could have been used to arrive at the human equivalent POD. This would have been more physiologically relevant than scaling according to BW ratios. Furthermore, mice epidermal thickness is 0.0104 mm vs 0.052 mm for humans which provides an epidermal thickness factor of 0.2 (Knafla et al., 2011).

The Knafla et al (2011) paper also proposed a skin cancer slope factor 3.5 per µg/cm²-day derived on a per unit skin surface area that can be used to estimate risks as a function of exposed surface area. The review is correct in stating that exposure to 0.01 ug/day over 10 cm² or 19,000 cm² could result in risk of a tumor; however, increased surface area can result in increased chemical flux and increased dose to target sites in skin.

For interspecies scaling of the oral slope factor, a different approach was used; TWAs were used here and not mentioned in the skin slope factor adjustment.

2.5.5. Uncertainties in the Derivation of the Dermal Slope factor

The review adequately addresses the uncertainty associated with the principal study and the supporting studies. The seven-fold difference in PODs is not alarming especially as they span 10 independent studies across male and female mice of at least 6 strains. Several of the concerns with the data sets such as low-dose extrapolation adds to the uncertainty in computation of slope factors from some studies. The solvent mixture (cyclohexane/acetone) used in the principal study does not reflect the real world dermal exposure for BaP, but it does provide an optimal exposure scenario. There is also some uncertainty in the extrapolation from animal skin exposure to human exposure. The statement on page 2-46, line 13-14, is not accurate; the toxicokinetics and toxicodynamics in mouse skin and human skin are not similar. Dermal absorption data in the literature demonstrated that there is an almost a 2-3 fold difference. Appendix D (page D-3) does not present a complete picture of BaP skin absorption in spite of the many publications available. For example, Ng et al (1992) and Sanders et al (1986) could be useful. Metabolism has also been shown to occur in basal cells which are located in the epidermis and not the dermis.

Question #5. Charge Question on Public Comments

Public Comment: *“Scientifically inappropriate to base human health risk assessment of hundreds of differing complex mixtures on the basis of one PAH, BaP.....” from Brian Magee (ACCCI, AFPM, AI, AAR, etc).”*

Response: The principal study for the slope factor determination (Sivak et al., 1997) did evaluate “real world” mixtures (fractions of asphalt roofing fumes) with BaP but it also evaluated BaP topically applied in a simple binary solvent system. There was a dose response with fume fractions with BaP concentrations. This could have been explained in the review to address this concern.

Public Comment: *“EPA review omitted entire literature on coal tar pharmaceutical products.”*
.... **“..there is little evidence that humans are at risk of developing skin cancer following dermal exposure to BaP”** . **“persuasive studies have grafted human skin onto mouse backs than then dosed with BaP...and these studies have repeatedly shown that the functioning human skin does not develop skin cancer as does the mouse skin beyond the margins of the graph”**: from *Brian Magee (ACCCI, AFPM, AI, AAR, etc)* . **“Pharmaceutical uses – was given short shift...failing to identify literature on exposures not associated with adverse effects”** from *Ann LeHuray, Pavement Coatings Technology Council.....*

Response: This was adequately answered in the review, For example: “Acute studies of coal tar 22 treated patients provide in vivo evidence of benzo[a]pyrene-specific genotoxicity (increased BPDE-23 DNA adducts) in human skin (Godschalk et al., 2001; Rojas et al., 2001; Zhang et al., 1990), an early 24 key event in the carcinogenic mode of action of benzo[a]pyrene (see Figure 1-6 of Section 1.1.5).”

However, the use of human-mouse skin xenographs in the review should be addressed. The review needs to address the weakness in these xenograph applications such as lack of positive controls, mouse life expectancy is 2 yrs and humans is 70 years, etc.

The presence of formulation additives in various pharmaceutical formulations (e.g., surfactant) are not often associated with increased dermal absorption but can more likely retain the chemical on the skin surface than cause epidermal penetration thereby limiting the effective dose to cause tumors.

Taken together, epidemiological studies that probe the link between skin cancer and topical exposure to pharmaceutical coal tar provided mixed results. In almost all cases the dose, duration and the appearance of skin tumor are not reported. The discovery of 50 mg/kg benzo[a]pyrene in cosmetic hair shampoos in Germany led the German government to ban coal tar products in 1992. Subsequently, German cosmetic manufacturers removed coal tar from their products. In 1997, the European Union placed refined coal tars on the list of substances that must be excluded from cosmetic products. There are epidemiological studies, also supported by anecdotal reports, that coal tar pharmaceutical can cause skin cancer.

- Sarto F, Zordan M, Tomanin R, et al. Chromosomal alterations in peripheral blood lymphocytes, urinary mutagenicity and excretion of polycyclic aromatic hydrocarbons in six psoriatic patients undergoing coal tar therapy. *Carcinogenesis*. 1989;10:329-334.
- Saperstein MD, Wheeler LA. Mutagenicity of coal tar preparations used in the treatment of psoriasis. *Toxicol Lett*. 1979;3:325-329.

A more recent 2015 study found no increase in skin or bladder cancer:

- Roelofzen JH, Aben KK, Van de Kerkhof PC, Van der Valk PG, Kiemeneij LA. [Dermatological exposure to coal tar and bladder cancer risk: a case-control study](#). *Urol Oncol*. 2015 Jan;33(1):20.e19-22. doi: 10.1016/j.urolonc.2013.12.006. Epub 2014 Mar 12.

FDA re-reviewed Coal Tar in 2001 in response to a citizens' petition. This review, including more recent epidemiology studies, confirmed Coal Tar as a Category I (safe and effective) OTC drug

ingredient (FDA 2001a; 2001b). The International Cosmetic Ingredient Dictionary and Handbook (Gottschalck and McEwen 2004) gives a function of Coal Tar in cosmetics as an antidandruff agent, industry reports of Coal Tar use in cosmetics are actually in OTC preparations at concentrations from 0.06% to 7% (CTFA 2002). There needs to be an estimate of what fraction of the 0.06-7% coal tar is BaP and furthermore, from that fraction, what proportion of BaP in that cosmetic formulation is available for transdermal diffusion. It is very likely that that fraction is very small and not of significance

Public Comment: *“the true and relevant dose of BaP in the skin is cumulative and increases over time. Erroneously using daily dose to derive a DSF results in meaningless DSF that is artificially high”*

Response: Human skin is a living membrane and BaP is transported by passive diffusion and not by membrane transporters which could be a reason for “depot” formation. In essence, a multiple dose regimen as often with other routes of exposure will result in steady state pharmacokinetics.

Public Comment: *USEPA (2013) needs to clearly state that any DSF that it may finalize in the future is focused on dermally absorbed dose and not applied dose. The studies used as the basis for the proposed DSF used soluble BaP in solvents that ensured that the BaP was completely absorbed into the skin. Real world exposures to BaP and other potentially carcinogenic PAHs are to complex mixtures and matrices that would impede the dermal absorption of the BaP.*

Response: This is a plausible observation. However, this will have to be applied to other routes of exposure where there is uncertainty in the dosimetry. However, the basis for the assessment is protection of the more sensitive population and scenarios/formulations/mixtures that will deliver BaP to the effector site.

Public Comment: *“Study Selection. Despite the weight of evidence that humans are not sensitive to chemically induced skin tumorigenesis as is the mouse skin and that PAHs build up in mouse skin after repeated dose administrations, USEPA (2013) has reviewed the mouse skin literature and chosen ten published papers as Key Studies. They exclude several studies by an arbitrary criterion: Study Duration. The excluded studies include:*

- *Levin et al. (1977)*
- *Nesnow et al. (1983)*

Response: EPA did not consider these studies because of a 1-time/week (Nesnow et al 1977) or 1-time every 2 weeks. This needs to be explained as the principal study (Sivak et al 1997) dosed twice weekly. (note application to shaved skin is not the same as clipped hair skin; the former increases permeability).

Dr. Annette Bunge

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

The following paper presents results related to vehicle effects and DNA adduct formation at remote sites following dermal exposure.

Booth, E.D., Loose, R.W., Watson, W.P. (1999). Effects of Solvent on DNA Adduct Formation in Skin and Lung of Cd1 Mice Exposed Cutaneously to Benzo(a)Pyrene. *Arch. Toxicol.* 73, 316-322.

The following papers include results of dermal absorption from BaP contaminated soils.

Abdel-Rahman, M. S.; Skowronski, G. A.; Turkall, R. M. Assessment of the Dermal Bioavailability of Soil-Aged Benzo(a)Pyrene. *Hum. Ecol. Risk Assess.* **2002**, 8, 429-441.

Yang, J. J.; Roy, T. A.; Krueger, A. J.; Neil, W.; Mackerer, C. R. Percutaneous Absorption of Benzo(a)Pyrene from Soils with and without Petroleum Crude Contamination. In *Petroleum Contaminated Soils*; Calabrese, E. J., Kostecki, P. T., Eds.; Lewis Publishers: Chelsea, MI, 1989; Vol. 2, pp 399-407.

Stroo, H. E.; Roy, T. A.; Liban, C. B.; Kreitinger, J. P. Dermal Bioavailability of Benzo[a]Pyrene on Lampblack: Implications for Risk Assessment. *Environ Toxicol Chem* **2005**, 24, 1568-1572.

The following paper provides a critical review of data quality and compares results from several studies of BaP absorption from contaminated soil studies

Spalt, E.W., Kissel, J.C., Shirai, J.H., Bunge, A.L. (2009). Dermal Absorption of Environmental Contaminants from Soil and Sediment: A Critical Review. *J. Expo. Sci. Environ. Epidemiol.* 19, 119-148, doi:10.1038/jes.2008.57.

3e. Dermal slope factor for cancer (section 2.5). The draft assessment proposes a dermal slope factor of 0.006 per ug/day based on skin tumors in mice. 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 scaling from mice to humans? Does the method for cross-species scaling (section 2.5.4 and appendix E) reflect the appropriate scientific considerations?

2.5.2. Dermal absorption and dosimetrics

2.5.4. Dermal Slope Factor Cross-Species Scaling

2.5.5. Uncertainties in the Derivation of the Dermal Slope Factor

RESPONSE:

Revisions to preliminary comments provided before the public meeting on April 15-17, 2015 are primarily related to clarifications of the absorbed versus exposed dermal dose and the exposed (or application) area versus contact area in the derivation and application of the dermal slope factor.

The challenge of IRIS assessments is to “identify potential adverse health effects and characterize exposure-response relationships” by integrating a wide range of scientific data (which were collected for diverse purposes, have varying quality and quantity, include contradictory results, and do not address many questions) and then extrapolating the results to human lifetime exposures with incomplete understanding of the fundamental biological processes. The aim is to protect (but not over protect) human health through scientifically reasonable and justifiable recommendations of exposure-response relationships with incomplete information. This review of the IRIS benzo[a]pyrene (BaP) document is mindful of the goal to make reasoned, scientific, “best guesses” from the present publically available information (with its limitations and flaws) to provide exposure-response relationships that could be used in risk assessments for identifying potential adverse health effects.

There is clear evidence that skin exposure to BaP causes skin cancer in mice, and most probably humans. Understanding the limitations of the available data, it is appropriate to develop a dermal slope factor (DSF) as plausible quantitative risk estimate for humans. EPA should consider revising the DSF from the present value 0.006 per ug/day by considering all rather than only one of the selected skin cancer bioassay studies in its derivation and after re-considering the chosen approach for scaling from mouse to human in the context of a hypothetical framework for skin cancer risk as discussed in the comments below.

Hypothetical framework for skin cancer risk should guide choices in dose metrics and scaling factor

Skin cancer risk from skin exposure to BaP depends on the levels of BaP metabolites in the cellular epidermis (i.e., the viable epidermis layer). Except when skin is damaged, BaP in the viable epidermis is the dose that absorbed into and through the outermost skin layer, the stratum corneum. Because the stratum corneum functions as a barrier that limits chemical exposure to tissues beneath it, the absorbed dose is less, often much less than the exposed dose (i.e., the BaP amount on the skin surface). In a typical human exposure, it is likely that a significant fraction of the exposed BaP dose will never be absorbed into the skin.

Although skin cancer risk depends on the absorbed dose, the dermal slope factor was derived from the exposed dose (also called the applied or administered dose) because the exposed dose is known, whereas the absorbed dose is not known unless the applied dose was absorbed completely. In the skin cancer bioassay studies, the dose is applied periodically without cleaning any remaining BaP residue from the skin surface prior to application. It is likely that nearly all of the applied BaP dose in the skin cancer studies is absorbed. The times between dose applications were long enough and the applied doses small enough relative to experimental rates of BaP skin absorption (e.g., Wester et al., 1990 reported 53% absorption into rhesus monkey

skin *in vivo* and 24% into human skin *in vitro* from a 24-h exposure to 0.5 µg BaP/cm²) that assuming ~100% absorption of the exposed dose is reasonable, provided that losses of the applied dose are minimal. Given this, it is likely that the derived values for point of departure and dermal slope factor are based on the absorbed dose. Therefore, human skin cancer risk estimated using the dermal slope factor should be calculated using the absorbed dose and not the exposed dose. That is, the lifetime average daily dose (LADD) should represent the absorbed dose calculated from the exposed dose and exposure scenario. This needs to be clearly explained in the document. (See additional comments below).

The description of the development of the dermal slope factor for cancer in humans would benefit from an outline of the steps to skin cancer as best they are currently understood or might be plausibly explained. Justification of choices in the dose-response analysis, methods for scaling the dermal slope factor from mice to humans, and adjustments for absorbed compared with exposed doses could then be related to assumptions about the mechanisms. This mechanistic framework might include:

1. Skin exposure to BaP (Controlling quantity or factors: exposed dose and scenario of exposure)
2. Absorption through the stratum corneum to reach the cancer forming tissue; i.e., the viable epidermis. (Controlling quantity or factors: absorbed dose and exposure scenario)
3. Local concentration of BaP in the cancer forming tissue; i.e., the viable epidermis (Controlling quantities or factors: cancer forming tissue mass, absorption to and clearance from cancer forming tissue)
4. Rates of metabolism and adduct formation within the cancer forming tissue (local concentration of BaP, BaP metabolite profile in mice and humans, metabolism rate, adduct formation rate)

Recognizing that my expertise is in dermal absorption and not in mechanisms of carcinogenesis, experts in carcinogenesis may have alternative recommendations for describing the steps related to cancer development.

Scaling the dermal slope factor derived from mice to humans

According to EPA guidance (U.S. E.P.A. 1992), the strategy when empirical data for developing a scaling factor are absent is to develop a “scientific rationale for a particular scaling factor by investigating the allometric variation of the biological features and processes that influence and underlie carcinogenic potency”. Therefore, lacking chemical specific data for oral exposures, the consensus for a typical or average chemical is that equal doses in units of mg/kg^{3/4}/day, when experienced daily for a full lifetime, will produce equal lifetime cancer risks across mammalian species (U.S. E.P.A., 1992). Scaling by the ³/₄ power of body weight is consistent with allometric variation of key physiological parameters across mammalian species (provided doses are low enough that saturation of enzyme activity is unlikely). For example daily intakes of food or water are approximately proportional to the ³/₄ power of body weight.

In the IRIS BaP document the selected scaling using body weight to the ³/₄ power is justified as follows (p. 2-44, lines 10-15):

“allometric scaling using body weight to the ³/₄ power was selected based on known species difference in dermal metabolism and penetration of benzo[a]pyrene. In vitro skin permeation was highest in the mouse, compared to rat, rabbit, and human, and was

enhanced by induction of CYP enzymes (Kao et al., 1985). Using this approach, rodents and humans exposed to the same daily dose of a carcinogen, adjusted for $BW^{3/4}$, would be expected to have equal lifetime risks of cancer.”

This same explanation is given in the EPA response to reviewers' comments (p. G-11, lines 48-49). Unfortunately, the IRIS BaP document failed to mention that the same paper (Kao et al. 1985) reported that BaP permeation rates through rabbit and rat skins were smaller (by a small amount) than through human skin (despite greater BaP metabolism in rabbit and rat skin), which seems to contradict the conclusion that risk scales with body weight to the $3/4$ power. As noted in Appendix E (p. E-112), Knafla et al., 2011 did review the biological features and processes that might affect extrapolating tumor potency in mice to humans. Based on their hypothesized mechanism of carcinogenesis and analysis of BaP metabolism to DNA adduct forming metabolites in human and mice, Knafla et al., 2011 chose to adjust the dermal slope factor in mice by a factor of 5, which represents the ratio of epidermal thickness in humans and mice. The Knafla et al., 2011 approach for scaling is mentioned in Appendix E to the IRIS BaP document, but no explanation was given for choosing to not use this approach. Instead, Appendix E presents four alternative approaches to cross-species scaling with minimal discussion on why scaling by body weight to the $3/4$ power was chosen instead of other alternatives.

I found the description of the four alternative approaches for cross-species (mouse to human) scaling of the dermal slope factor unnecessarily confusing. Since the scaling is only from mouse to human, it would be simpler and more consistent in its application to the four approaches (and less confusing) to start with the dermal scaling factor in the mouse (i.e., $DSF_M = 1.7 (\mu\text{g}/\text{day})^{-1}$), and then adjust it by the appropriate human to mouse ratio to obtain the dermal slope factor in humans (DSF_H) for each approach as follows:

Approach 1. No interspecies adjustment: $DSF_H = DSF_M / 1$

$DSF_H = 1.7 (\mu\text{g}/\text{day})^{-1}$

Approach 2. Scaling by surface area: $DSF_H = DSF_M / (SA_H / SA_M) = DSF_M / (19,000 \text{ cm}^2 / 100 \text{ cm}^2) = (1.7 (\mu\text{g}/\text{day})^{-1}) / 190$

$DSF_H = 0.0089 (\mu\text{g}/\text{day})^{-1}$

Approach 3. Scaling by body weight: $DSF_H = DSF_M / (BW_H / BW_M) = DSF_M / (70 \text{ kg} / 0.035 \text{ kg}) = (1.7 (\mu\text{g}/\text{day})^{-1}) / 2000$

$DSF_H = 0.00085 (\mu\text{g}/\text{day})^{-1}$

Approach 4. Scaling by body weight to the $3/4$ power: $DSF_H = DSF_M / (BW_H / BW_M)^{0.75} = DSF_M / (70 \text{ kg} / 0.035 \text{ kg})^{0.75} = (1.7 (\mu\text{g}/\text{day})^{-1}) / 2000^{0.75} = (1.7 (\mu\text{g}/\text{day})^{-1}) / 300$

$DSF_H = 0.0057 (\mu\text{g}/\text{day})^{-1}$

In the above calculations, SA_j = total skin surface area, BW_j = body weight, and the subscript j designates that the quantity is specified for mouse (M) or human (H). Cancer risk predicted by all four approaches is then calculated by multiplying the DSF_H by the same lifetime averaged dermal dose (LADD), which has units of $\mu\text{g}/\text{day}$. For the example calculations listed in Table E-25, LADD is $0.0004 \mu\text{g}/\text{day}$. The DSF for Approach 4 is reported in Appendix E as listed above. However, the DSF for Approaches 2 and 3 are listed as $170 (\mu\text{g}/\text{cm}^2 \text{ day})^{-1}$ and as $0.058 (\mu\text{g}/\text{kg}$

day)⁻¹, which requires that the LADD values used for calculating risk be divided by 19,000 cm² surface area for Approach 2 and by 70 kg for Approach 3. This is confusing and these adjustments to LADD are not described clearly.

It is recommended that Table E-25 be revised as follows: (1) the dose metric column should be deleted; (2) the dermal slope factor numbers should be listed as given above (i.e., 1.7, 0.0089, 0.00085, 0.0057 for Approaches 1 through 4, respectively); these numbers all have the same units, which can be listed in the column heading; (3) the reference to “nominal exposure” should be called the “lifetime averaged dermal dose” (LADD) instead to be consistent with the terminology elsewhere, (4) the LADD of 0.0004 µg/day is based on exposed dose; it should be based on absorbed dose.

Assumptions of the approaches for scaling the dermal slope factor from mice to humans

I found the Assumptions listed in Table E-25 for the four approaches confusing. First, in all four approaches, the risk in a given species is assumed to depend on only the overall dose and not the dose per unit area or body weight. For a given LADD (e.g., 0.0004 µg/day listed in Table E-25), cancer risk is different for the different approaches. However, the cancer risk in a human is the same whether 0.0004 µg/day is applied to a skin area of 1 cm² (dose per area = 0.0004 µg/cm²-day) or 19,000 cm² (dose per area = 2.1E-8 µg/cm²-day). Thus, I believe that the last sentence of the Assumptions listed in Table E-25 for Approach 2:

“This approach implies that risk does not increase with area exposed as long as dose per area remains constant.”

is incorrect. If dose per exposed area is constant, then increasing the exposed area increases the LADD (which is calculated based on total skin area) and therefore, the cancer risk.

Other statements in Table E-25 were confusing, partly because it was not always clear whether the word “area” referred to the total exposed area across which the vehicle containing the BaP is spread or to the area with direct contact to BaP (or to the vehicle containing the BaP).

Depending on the scenario (i.e., type of vehicle and exposure situation), the contact area can be less than or equal to the exposed area. In the case of neat BaP deposited onto skin in a volatile solvent (as in most of the mouse skin cancer bioassay studies), the solvent evaporates leaving small “piles” of solid BaP that contact a much smaller area than the total area exposed originally to the solvent containing the BaP. I will refer to this situation as Scenario 1. Scenario 1 also applies when skin is exposed to BaP contaminated soils, in which the typical soil adherence factors (i.e., the mass of soil per skin area) are much smaller than the soil mass required to cover the exposed area with a soil monolayer. For Scenario 1, the skin area with direct contact to BaP does not change if the equal mass per day dosing is spread across a larger exposed area.

Skin exposures to BaP dissolved in non-volatile oils or coal tar ointments can completely cover the exposed area. In this scenario, which I will call Scenario 2, the skin area in contact with BaP increases if the vehicle containing the BaP is spread across a larger exposed area (as long as the area is not increased beyond the capability of the vehicle to spread over the entire area). For Scenario 2, the skin area in contact with BaP does change if the equal mass per day dosing is spread across a larger area. Whether the BaP concentration in the non-volatile liquid or semi-liquid vehicle changes when an equal mass per day dosing is spread across a larger area depends

on whether the total volume of the vehicle is held constant (BaP concentration in the vehicle does not change, but the thickness of the vehicle film decreases when spread over a larger area) or the volume of vehicle per skin area is held constant (BaP concentration decreases proportionally with the volume increase while the thickness of the vehicle film does not change with increasing volume).

The assumptions of the approaches listed in Table E-25 should be written with awareness of these two different scenarios, or specify that they apply to one of the scenarios.

The first 3 sentences of the Assumptions listed in Table E-25 for Approach 3:

“The skin is an organ with thickness and volume; benzo[a]pyrene is distributed within this volume of skin. Cancer risk is proportional to the concentration of benzo[a]pyrene in the exposed volume of skin. Equal mass per day ($\mu\text{g}/\text{d}$), if distributed within equal fractions of total body skin will have similar cancer risks. That is, whole-body lifetime exposure (e.g., 5%-of-the-body lifetime exposure) at the same loading rate ($\mu\text{g}/\text{cm}^2\text{-d}$) gives similar cancer risks across species.”

which expand on the statement:

“This approach assumes that risk is proportional to dose expressed as mass per kg body weight per day.”

could be stated more simply as: (1) equal average BaP skin concentrations, which correspond to mass per day ($\mu\text{g}/\text{d}$) per total skin mass, will have similar lifetime cancer risk, and (2) the mass of skin is assumed to be approximately the same fraction of the body weight in all species.

Scaling approaches based on mechanism may provide valuable insights

In my mind, scaling factors for skin cancer arising from BaP in the viable epidermis target tissue should be bounded by the extremes of (1) total skin area (i.e., 19,000 cm^2 in humans compared with 100 cm^2 in mice), and (2) skin area with BaP contact (i.e., the same contact area for the same applied dose in humans and mice). For either of these, scaling could depend on the skin area ratio of the viable epidermis (target tissue), the volume ratio of the viable epidermis, and the (3) metabolic rate ratio of the viable epidermis. The volume ratio would be equal to the product of the surface area ratio and the ratio of the viable epidermis thickness. The metabolic rate ratio would be equal to the product of the volume ratio and the ratio of the metabolic rates per tissue volume. I have not included effects related to the stratum corneum in the interspecies scaling factor, such as permeation rates, because these factors would affect the estimate of the absorbed dose compared with the exposed dose, but not the cancer risk for the same absorbed dose.

The estimated human to mouse scaling factors for these cases are listed in the following table, assuming the following input parameters:

- Total skin surface area for human (19,000 cm^2) and mouse (100 cm^2)
- Ratio of the viable epidermis thickness in human to mouse is 5 (from Knafla et al., 2011)
- Body weight for human (70 kg) and mouse (0.035 kg)

Ratio of indicated quantity	Definition	Estimate based on <u>total skin area</u> in mice & humans	Estimate based on <u>contact area</u> (assumed to be the same in mice & humans for same BaP mass/d)

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Area	A_H/A_M	$19,000/100 = 190$	1
Volume	$V_H/V_M = (A_H/A_M)(L_H/L_M)$	$190 \times 5 = 950$	$1 \times 5 = 5$
Metabolic rate	$(V_H/V_M)(r_{met,H}/r_{met,M})$	$950 \times (r_{met,H}/r_{met,M})$ perhaps this is equal to: $(BW_H/BW_M)^{3/4} = (70/0.035)^{3/4}$ $= 2000^{3/4} = 300$ from the above we might estimate $r_{met,H}/r_{met,M}$ as: $r_{met,H}/r_{met,M} = 950/300 = 1/3.2$	$5 \times (r_{met,H}/r_{met,M})$ If $r_{met,H}/r_{met,M} = 1/3.2$, then $5 / 3.2 = 1.6$

where A is the area, V is the volume, L is the thickness, and r_{met} is the metabolic rate per volume of tissue for the viable epidermis (the target tissue) for humans or mice as indicated by subscript H or M, respectively. BW is the body weight of humans or mice as indicated by subscript H or M, respectively. It is suggested in the above analysis that the metabolic rate ratio of the viable epidermis might be represented by the allometric scaling of body weight ratio raised to the $3/4$ power. This would suggest that the Approach 4 allometric adjustment used to derive the proposed DSF of 0.006 per $\mu\text{g}/\text{day}$ is based on total skin area combined with relative metabolic rates in the viable epidermis. Other information about the metabolic rates in the viable epidermis tissue might be found in the published literature and used in the metabolic rate ratio.

Note that the four mouse-to-human scaling approaches considered in the IRIS BaP document are included in the table. Scaling by the contact area ratio corresponds to Approach 1 (no scaling). The area ratio based on total skin area is Approach 2. The volume ratio based on total skin area is essentially Approach 3, except that the volume ratio in the table above has been estimated from the skin area ratio and the ratio of viable epidermis thickness (which gives a scaling factor of 950) rather than assuming the skin volume fraction is the same for all species (which gives a scaling factor of 2000 calculated from the body weight ratio). The metabolic rate ratio based on the total skin area is expected to be essentially Approach 4. The volume ratio based on contact area is the approach used by Knafla et al., 2011.

The skin area exposed to BaP is not the total skin area in either mice or humans. In humans, the exposed skin surface area could vary with the exposure scenario (e.g., the skin area exposed to contaminated soils for a landscaper would be different than for an office worker). In the mice skin cancer bioassay experiments, the exposed area usually was not specified, but it was probably 5 to 10 cm^2 based on a rough estimate of interscapular area of the adult mouse back, which was the application site. For example, Knafla et al., 2011 reported 6 cm^2 was used in the study by Nesnow et al., 1983. An exposed area of 5 to 10 cm^2 represents approximately 5 to 10% of the total skin area in mice. Note that the mouse-to-human scaling factors based on total skin area are also consistent with applying the same BaP mass per day to the same fraction of the total area on mice and humans (e.g., 10 cm^2 exposed area on mice and 1900 cm^2 exposed area on humans, which is 10% of the total skin area). This was also recognized in the IRIS BaP document (see the description of Approach 2 scaling described in section E.2.3 and Table E-25 of the Supplemental Information). Therefore, it would seem that scaling factors based on the total skin area should include an adjustment (ADJ_{SA}) for different skin area fractions of exposed

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

skin in the mouse and human ($ADJ_{SA} = (SA_H/A_H) / (SA_M/A_M)$ where SA is the exposed skin area). Thus,

$$\text{Human Cancer Risk} = DSF_H \times ADJ_{SA} \times LADD = (DSF_M/\text{ScalingFactor}) \times ADJ_{SA} \times LADD$$

A human skin exposure to BaP on an area of 950 cm² (the area of the example cancer risk calculation in Table E-25) might be approximately the same area fraction as the exposed area in the mouse skin cancer bioassays (i.e., 5% area), so an adjustment for different area fractions would not be needed (i.e., $ADJ_{SA} \approx 1$). However, a human exposure to an area of 5700 cm² (the area of the example cancer risk calculation on page G-16 of the Supplemental Information) is approximately 30% of the total skin area of a human, which would correspond to $ADJ_{SA} \approx 6$ if the exposed area fraction in mice is estimated to be 5% (or 3 if the exposed area fraction in mice is estimated to be 10%). In the hypothetical framework outlined in the table above, ADJ_{SA} should be applied to any scaling factor derived from total skin area, which includes the metabolic rate ratio that is essentially Approach 4 in the IRIS BaP document.

As used in the calculations shown in the above table, the contact area would correspond to a unit of area for each unit of BaP applied in each day. That is, a unit of BaP mass that contacts the same skin location on two days is considered to have the same cancer risk as if the same unit mass of BaP contacts different skin locations on two days. Scaling factors based on the skin contact area would be independent of the exposure scenario.

I have not resolved in my mind which of these scenarios is most plausible (or if there are other more plausible scenarios that I have not thought of yet perhaps related to DNA adduct formation). However, reasoning based on differences in biology of the target tissue and the exposure scenarios should be considered in identifying the chosen scaling factor, concluding that the chosen scaling factor is (or is not) independent of the exposure scenario (note that EPA presently assumes that the scaling factor is independent of the exposure scenario), and estimating the magnitude of the uncertainty in the chosen scaling factor. For example, assuming that the metabolic rates in the target tissue determine the relative cancer risk in humans and mice (which might be approximately estimated for the skin area by the body weight ratio raised to the $\frac{3}{4}$ power), then the human-to-mouse scaling factor is expected to be between 2 (based on contact area) and 300 (based on total skin area without an adjustment for different area fractions exposed). Additional insights might be brought to the problem to reduce this range. These might include an assessment of reasonableness of the resulting DSF.

Example estimates of skin cancer risk for the different scaling approaches

The numbers listed as “Risk at nominal exposure” in the right hand column of Table E-25 were calculated using 0.0004 µg/day as the “nominal exposure”. The derivation of 0.0004 µg/day described in footnote “a” of Table E-25 is inadequate. Users of this document should be able to understand readily the calculation of this number; all numbers used in the calculation of 0.004 µg/day should be listed in footnote “a”. Also, as stated above, “nominal exposure” should be called the “nominal lifetime averaged dermal dose” or “nominal LADD”. Most important, the description of how to calculate the LADD for an example exposure scenario provides an opportunity to clearly explain that, to be consistent with the derivation of the DSF, the LADD should be calculated using the absorbed dose that is estimated from the exposed dose. The

equation for estimating the example LADD for Table E-25 needs to be provided along with specification of all the parameter values (similar to the list of parameter values provided for an adult on p. G-16), including a parameter for adjusting the exposed dose to give the absorbed dose.

EPA should consider dividing the equation for LADD into two equations that separate the lifetime average dose calculation from the daily absorbed dose calculation (similar to the calculations in U.S. EPA, 2004). In this scheme, $LADD_{\text{dermal}}$ would be calculated using the average dose absorbed each day from skin exposure to BaP (DA_{day}).

$$LADD_{\text{dermal}} = \frac{DA_{\text{day}} \times ED \times EF}{AT}$$

where ED, EF and AT are defined as listed on p. G-14 of the supplemental information document. When written this way, the calculation for LADD is the same for exposures to all types of environmental media. The calculation of DA_{day} would be defined by a separate equation, which will be different for skin exposure to BaP in different environmental media. For exposure to BaP on soil, the second equation would be:

$$DA_{\text{day}} = C_{\text{soil}} \times CF \times SA \times AF \times ABS_{\text{dermal}}$$

where C_{soil} , CF, SA, and AF are defined as listed on p. G-14 of the supplemental information document, and ABS_{dermal} is the fraction of the exposed dose that absorbs. The value used for ABS_{dermal} should be measured from or estimated for absorption of BaP into skin (i.e., the amount of BaP in the skin at the end of the exposure period plus the amount that has penetrated through the skin during the exposure period) from a monolayer of soil applied to skin.

As best as I can determine, the “nominal exposure” of 0.0004 $\mu\text{g}/\text{day}$ in Table E-25 was calculated using $C_{\text{soil}} = 100$ ppb (as listed in Table E-25), SA (surface area exposed, as listed on p. E-114, line 9) = 950 cm^2 , AF (soil adherence factor) = 10 $\mu\text{g}/\text{cm}^2\text{-d}$, ED (exposure duration, I think from (U.S. E.P.A. 2004)) = 30 years, EF (exposure frequency) = 350 days/yr, and AT (lifetime averaging time) = 365 days/yr*70 years = 25550 days. Thus, this calculation differs from that shown on p. G-16, which is for a SA = 5700 cm^2 , an exposure duration of 9 years and includes the soil-to-skin transfer coefficient (K_{soil}) = 0.25. Note that K_{soil} is not the same as ABS_{dermal} . The parameter $K_{\text{soil}} = 0.25$ in the example calculations shown on pages G-14 to G-16 was the ratio of the percentage BaP absorbed from soil (13%) to that from acetone (51%) as reported by Wester et al., 1990 for *in vivo* rhesus monkeys. It was used to estimate an effective exposed dose (i.e., to adjust the total mass of BaP on the soil to the fraction that has contact with skin). To be consistent with the development of the dermal slope factor, the LADD should be calculated using ABS_{dermal} to estimate the absorbed dose from soils rather than K_{soil} to estimate an effective exposed dose. As an example, $ABS_{\text{dermal}} = 0.13$ might be used consistent with the recommendations from (U.S. EPA, 2004). Additional comments are provided below on choosing values of ABS_{dermal} for BaP on contaminated soils.

Finally, it is recommended that example calculations of estimated cancer risk from soil exposures provided in different places in the document be consistent with each other. That is currently not the case as K_{soil} is used in the examples provided on pages G-14 to G-16, but not in the examples listed in Table E-25. The LADD in all examples should represent the absorbed dose.

Relating BaP cancer risk derived from solvent deposited BaP to contaminated soils

The Preface to the IRIS BaP document (pages xii-xiii) explains that the IRIS Program's first dermal slope factor was motivated by the "Agency's need to estimate the potential for skin cancer from dermal exposure (U.S. EPA, 2004), especially in children exposed to contaminated soil...". Because cancer bioassays from skin contact to BaP contaminated soil are not available, the dermal slope factor was derived for BaP delivered to skin using volatile solvent solution (typically acetone or toluene) that evaporated leaving BaP in direct contact with skin. A similar BaP dose on soil will have a reduced cancer potency compared to the solvent delivered BaP. Moreover, in actual human exposures, some of the exposed dose is removed during daily activities such as bathing, which did not occur in the cancer bioassay experiments.

Although skin exposures to BaP contaminated soil motivated the slope factor derivation, the IRIS BaP document provides minimal information on how to use the DSF for estimating cancer risk from exposure to contaminated soils. While there are no cancer studies of skin contact to BaP contaminated soil, there are dermal absorption studies from both soil and volatile solvents, which could be considered. One study (Wester et al. 1990) is mentioned on pages G-12 (lines 35-43) and G-16 in the EPA response to comments, along with the soil-to-skin transfer coefficient (K_{soil}) of 0.25. As discussed above, the estimate of LADD should be based on the absorbed dose (estimated using ABS_{dermal}) rather than an adjusted exposed dose (estimated in the examples shown on pages G-14 to G-16 using K_{soil} calculated from dermal absorption measurements from soil compared to acetone). Although dermal absorption measurements of BaP deposited onto skin from volatile solvents like acetone are not needed for comparison with dermal absorption from soils, they are useful for establishing that most of the BaP applied to skin in the skin cancer bioassay studies in mice probably absorbed. Estimates for ABS_{dermal} should be based on studies of BaP absorption into skin from contaminated soils.

Skin cancer risk from skin exposure BaP depends on the levels of BaP metabolites in the cellular epidermis (i.e., the viable epidermis layer). Therefore, except when skin is damaged, BaP in the viable epidermis is the dose that absorbed into and through the outermost skin layer, the stratum corneum. This will be essentially the same as the systemically absorbed dose. In the *in vitro* experiment, the systemically absorbed dose will be represented by the amount that appears in the receptor fluid during the experiment plus the amount present in the washed skin (i.e., after excess, non-absorbed BaP has been cleaned from the skin surface) at the end of the experiment. There are several published studies of dermal absorption from BaP contaminated soils (Abdel-Rahman et al. 2002; Moody et al. 1995; Moody et al. 2007a; Roy and Singh 2001; Roy et al. 1992; Roy et al. 1998; Stroo et al. 2005; Stroo et al. 2000; Turkall et al. 2008; Wester et al. 1990; Yang et al. 1989a; Yang et al. 1989b). The results from some of these studies do not reliably represent dermal absorption from soils; e.g., the BaP contaminated soils contained water or solvents as discussed in Spalt et al. (Spalt et al. 2009). We have worked over the past few years on a thorough review of published studies of BaP absorption from skin exposure to contaminated soils, which we hope to submit for publication in the near future. This is a complicated literature and it is probably unrealistic to expect that a review could be added to the IRIS BaP document. In lieu of this, EPA should list references for the BaP contaminated soil studies, indicate that a full review has not been done (and is needed), and explain that the absorption estimate used in the example calculations is not an endorsement of this as the best value. Some of the factors that

should be considered for a scientifically supportable estimate of skin absorption could be mentioned. For example, the mass of soil applied in most soil studies was large enough that only a small fraction of the soil was in contact with the skin; i.e., the soil mass was larger than the mass required to cover the skin with a single tightly packed layer (a monolayer) of soil. A complication is that the mass of soil in a monolayer depends on the particle sizes of the soil (see Exhibit C-4 in (U.S. E.P.A. 2004)), which have varied drastically among the studies. At 40 mg of soil/cm², the soil load in the Wester et al., 1990 experiments was larger than other soil studies (Spalt et al. 2009). Despite this, it is likely that the skin was covered with approximately one layer of particles because the particle size fraction they used was large (i.e., 180 - 320 µm) (Spalt et al. 2009). This makes direct comparison of their results in terms of percent absorption appropriate; this would not have been the case if there had been multiple soil layers in their experiments.

Note also that Wester et al. (1990) reported results from studies in human skin *in vitro* as well as Rhesus monkeys *in vivo*. The BaP document only mentions the *in vivo* results (13% from soil on pages G-12 and G-16). Based on an examination of the experimental protocol used in the *in vivo* soil measurements, Spalt et al. (Spalt et al. 2009) questioned the reliability of the 13% absorption value reported by Wester et al., 1990. Certainly, in comparison with the other literature of BaP absorption from soils into human skin, the *in vivo* result from Wester et al., 1990 is significantly higher and much higher than their *in vitro* soil measurement (1.4 % BaP in the skin and receptor fluid after 24 h). In contrast, the *in vitro* and *in vivo* measurements from acetone deposited BaP are more similar to each other than the soil results (only a factor of two different – 51% versus 24% when deposited from acetone compared with a factor of almost 10 different – 13% versus 1.4% from soil). Note that the Knafla et al., 2011 also considered soil studies by Moody et al., 2007 and Abdel-Rahman et al., 2002 in their examination of the relative absorption factor. It should be noted, however, that absorption measurements for BaP reported by Moody et al., 2007 and Abdel-Rahman et al., 2002 contaminated soils may not represent dermal absorption from soils. In the studies by Moody et al., 2007, soils were suspended in water, and thus, the results are pertinent to absorption from water in contact with soil but not to skin exposures to contaminated soil. In the studies of freshly contaminated soil by Abdel-Rahman et al., 2002 it appears that BaP solvent was added to the soil after it was placed on the skin (Spalt et al. 2009).

Dosemetrics in the skin cancer bioassays and estimates of human skin cancer risk

The dermal slope factor (DSF) was derived using the average continuous daily exposure to BaP mass for skin cancer bioassay protocols with two or three applications per week to mice. After scaling from mouse-to-human, it is intended that the DSF be used with estimates of “constant daily lifetime exposure” (quote from line 17, page 2-41) in humans. Therefore, in both its derivation from mice skin cancer bioassays and its application for estimating human cancer risk from skin exposures, the DSF is based on the dose as mass of BaP and not the mass of BaP per unit area of skin. Thus, a mass of BaP contacting a smaller area of exposed skin is assumed to have the same risk of skin tumors as the same mass of BaP applied to a larger skin area. In most of the skin cancer bioassay studies, neat BaP was applied to skin in volatile solvent that evaporated soon after application. From the perspective of dermal absorption, the mass of BaP that absorbs into and through the skin depends on the total BaP mass and not the BaP mass per area for doses that are smaller than the amount of BaP that would be required to cover an

exposed skin area with a complete film of BaP. Thus, at low doses, piles of neat BaP solid will cover a fraction of the skin area to which BaP in the volatile solvent was applied. Therefore, the skin area in direct contact with neat BaP does not change if the same mass of BaP is applied to a larger area of skin. Also, increasing the amount of BaP applied to the same area proportionally increases the skin area fraction covered with neat BaP. The same is true for BaP on contaminated soils that only partially covers the exposed area (see Roy and Singh, 2001). BaP must absorb into and through the stratum corneum to reach the target tissue for skin cancer (the viable epidermis), which, like the stratum corneum, is not vascularized. Therefore, BaP contact with the viable epidermis is likely to coincide with the area of BaP contact with the stratum corneum surface. Given this, I agree with the conclusion of the IRIS BaP document that the surface area over which the BaP dose is applied does not need to be considered.

Related to this, it is not clear to this reviewer why studies that applied BaP once/week (e.g., Nesnow et al., 1983) or once every 2 weeks (e.g., Levin et al., 1977) are "less useful for extrapolating to daily human exposure" (see page D-62, lines 8-10) than studies that applied BaP 2-times or more per week. If the results of applying BaP once/week differ from applications of 2-times or more per week, then continuous daily exposure, which has been assumed in the analysis for the dermal slope factor is inappropriate; i.e., there would be data indicating that dose-rate effects cannot be ignored (see lines 12-13, p. 2-41).

Note that the quotation from line 17 of page 2-41 should be changed to say that it is intended for the DSF to be used with estimates of dermal absorption from "constant daily lifetime exposure".

Comments about vehicle effects

Vehicle effects in the skin cancer bioassay studies were probably minimized because the applied dose was almost completely absorbed. (This may not be the case for skin cancer bioassay studies of other chemicals). If only a fraction of the applied BaP was absorbed, then skin absorption and tumor development would likely have varied with vehicle. As an example, in studies by Grimmer et al. (1983, 1984), BaP was applied using solvent solution of 1:3 v:v dimethyl sulfoxide:acetone. Acetone is more volatile than dimethyl sulfoxide (DMSO) and would rapidly evaporate leaving BaP in a DMSO solution (at 25°C, vapor pressure of acetone is 231 mm Hg compared with 0.61 mm Hg for DMSO). *In vitro* diffusion cell studies of n-alcohols (methanol, butanol and octanol) at 1 % concentrations in DMSO:water solutions varying from zero to 100% DMSO have shown that skin is damaged irreversibly at DMSO concentration of 50% or larger after less than 3 h of exposure (Kurihara-Bergstrom et al. 1986). (Similar findings have been observed by other authors for other permeants as well.) Thus, skin damage that might have promoted skin penetration and perhaps tumor development may have occurred in the Grimmer et al. studies that would not have occurred in studies that used a vehicle of only volatile solvent(s). The similarity of the results from Grimmer et al. and the other studies listed in Table 2-11 is consistent with the expectation that most of the applied BaP absorbed into the skin.

An example of vehicle effects was observed in mice treated with a single 25- μ L application of the same BaP concentrations (ranging from 0.4 to 280 μ g/mL) in n-dodecane or tetrahydrofuran for 24 h (which was too short for all of the BaP to absorb). Fewer DNA-adducts were observed from the lipophilic and less volatile n-dodecane (vapor pressure = 0.135 mm Hg at 25°C) vehicle (0.067 to 3.5 fmol adducts/ μ g DNA) compared with the more polar and more volatile

tetrahydrofuran (vapor pressure = 162 mm Hg) vehicle (0.089 to 16.9 fmol adducts/ug DNA) (Booth et al. 1999). The difference may be that the BaP remained dissolved at less than its saturation concentration in n-dodecane for much of the exposure time, while the BaP in tetrahydrofuran had direct skin contact for much of the exposure time because tetrahydrofuran evaporated soon after application.

Extrapolation of cancer risk estimated from experiments that deposited BaP onto skin from a volatile solvent to BaP in other vehicles (e.g., solutions that do not evaporate or PAH containing ointments) must consider how the vehicle affects the driving force for BaP transfer into and through the skin (i.e., the thermodynamic activity) as well how vehicle components could affect the skin (e.g., enhance or retard absorption). Such considerations would need to occur in the estimates of the dermally absorbed dose, which is the correct dose to be used with the DSF.

Other specific comments (in no particular order):

The statement on p. D-3, lines 6-10:

“Studies of benzo[a]pyrene metabolites or DNA adducts measured in humans exposed dermally to benzo[a]pyrene-containing PAH mixtures demonstrate that benzo[a]pyrene is absorbed dermally. One study of dermal absorption in volunteers found absorption rate constants ranging from 0.036 to 0.135/hour over a 45-minute exposure, suggesting that 20–56% of the dose would be absorbed within 6 hours (VanRooij et al., 1993).”

is potentially misleading and should be revised. The absorption rate constants, which appear to be taken from Table 2 of VanRooij et al. (1993), are not for BaP as implied; they are PAHs with four or more fused ring PAHs after application of coal tar ointment. These absorption rates depend on concentrations of the various PAHs in the coal tar ointment as well as the amount applied, which are both unknown. Moreover, the assumption that the absorption rates for PAHs with four or more fused rings will represent BaP is not stated or justified.

I found two examples citations of incorrect citations. The first of these was also mentioned by Brian Magee in comments he submitted (p. 81) in November 2013 (on behalf of a consortium of trade groups). He noted that the Grimmer et al., 1984 citation for the mouse skin painting study is incorrect. He suggested that the correct citation is to Grimmer et al., 1985. However, there is also a different 1984 paper by Grimmer et al. that also contains mouse skin painting data. I have not examined these two papers, which are cited below, to determine which reference is the correct one. All of the numerous citations to this reference in the main IRIS BaP document and the Supplemental Information need to be checked.

G. Grimmer, H. Brune, R. Deutsch-Wenzel, G. Dettbarn, J. Misfeld, U. Abel, J. Timm (1984). The contribution of polycyclic aromatic hydrocarbons to the carcinogenic impact of emission condensate from coal-fired residential furnaces evaluated by topical application to the skin of mice. *Cancer Lett*, 23:167-176.

G. Grimmer, H. Brune, R. Deutsch-Wenzel, G. Dettbarn, J. Misfeld, U. Abel, J. Timm (1985). The contribution of polycyclic aromatic hydrocarbon fractions with different boiling ranges to the carcinogenic impact of emission condensate from coal fired residential furnaces as evaluated by topical application to the skin of mice, *Cancer letters*, 28, 203-211.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Godschalk et al. 1998a should be Godschalk et al., 1998b and the reference to Godschalk et al., 1998b should be 1998a.

There may be other citation errors. It is strongly recommended that the main and supporting information documents be thoroughly and carefully check all citations.

EPA states on p. G-12 (lines 22-24): the assumption of this dose metric is that risk at low doses of BaP is dependent on absolute dermal dose and not dose per unit of skin". Note that this should say "unit of skin area". Also, this idea is stated more clearly at the bottom of p. E-113 and continuing on to p. E-114:

"Risk at low doses of benzo[a]pyrene is dependent on absolute dermal dose and not dose per unit of skin [add the word "area"], meaning that a higher exposure concentration of benzo[a]pyrene contacting a smaller area of exposed skin could carry the same risk of skin tumors as a lower exposure concentration of benzo[a]pyrene that contacts a larger area of skin."

Consider revising the text on p. G-12 to make it more similar to the clearer description on p. E-114.

Regarding the statement on p. 2-42, lines 16-21:

"Dermal slope factors calculated from the supporting studies (Sivak et al., 1997; Grimmer et al., 1984; Habs et al., 1984; Grimmer et al., 1983; Habs et al., 1980; Schmähl et al., 1977; Schmidt et al., 1973; Roe et al., 1970; Poel, 1963, 1959) using the multistage model and linear extrapolation from the BMDL₁₀ values ranged from 0.25 to 1.8 per µg/day, a roughly sevenfold range (Table 2-11). Values ranged from 0.9 to 1.7 per µg/day for male mice, and from 0.25 to 0.67 per µg/day for female mice."

The numbers in the above statement do not seem to correspond to the numbers listed in Table 2-11. In my reading of Table 2-11, it seems that for males the range is 1.3 to 1.7 (only 2 studies) and for females the range is 0.25 to 1.8 (8 studies).

In Table 2-11, I believe that POD = BMDL should be POD = BMDL₁₀ to agree with the text describing Table 2-11. Additional information on the meaning of "no characterization of exposure duration" would be helpful. Is it the same for all the studies identified as having "no characterization of exposure duration"?

The IRIS BaP document explains on p. 2-45 that the dermal slope factor has been developed for local effect and not systemic cancer risk. It cites Godschalk et al. (1998a), which should be (1998b), as a source of information suggesting that BaP metabolites can enter the systemic circulation following dermal exposure in humans. Booth et al (1999) might also be cited as evidence that dermal exposure could cause systemic effects. They observed DNA adduct formation in the lungs of mice exposed to a single dermal dose of BaP.

The discussion on p. 2-45 continues by explaining that tumors have not been found at distal sites in lifetime skin cancer bioassays that included pathological examination of other organs. This is an expected observation because the concentration of BaP in skin (i.e., the portal of entry and a smaller tissue volume) will be larger than the systemic concentration, consisting of large tissue volume in which BaP is diluted. The explanation of skin binding and reactive metabolites in the

skin (lines 10-17) is unnecessary and irrelevant. Experimentally, a significant amount of BaP is found in the skin at the end of the dermal absorption experiments (typically lasting 24-h or less) because it takes time for the BaP to absorb into and then transfer through the skin to enter the systemic circulation. Had the experiment continued for a longer time, almost all of the BaP would have been excreted or found in the systemic circulation (in the *in vivo* study) or in the receptor fluid (in the *in vitro* study if the time required were not too long to maintain skin integrity).

Adding a footnote to Tables E-20, E-21 and E-22 (which appear on pages E-79 and E-80) that explains how the average daily dose was adjusted to lifetime averaged dose would help readers who may not remember the explanation provided in the text on p. E-75.

p. G-12 (lines 22-26) states:

“The assumption of this dose metric is that risk at low doses of benzo[a]pyrene is dependent on absolute dermal dose and not dose per unit of skin, meaning that a higher exposure concentration of benzo[a]pyrene contacting a smaller area of exposed skin could carry the same risk of skin tumors as a lower exposure concentration of benzo[a]pyrene that contacts a larger area of skin.”

This statement confuses concentration (which is mass per volume of the environmental media) with mass per area and contact area versus exposed area. Perhaps the intended thought was that the mass of BaP dissolved into a smaller volume of solvent solution (i.e., a higher concentration in the solvent) will spread across a smaller exposed skin area than the same mass of BaP dissolved into a larger volume of solvent solution (i.e., a lower concentration in the solvent), which will spread across a larger exposed area of skin. In most of the BaP skin cancer bioassay studies, the solvent was volatile and pure solid BaP was deposited directly onto the skin once the solvent evaporated. In this scenario, the skin area that has direct contact with the deposited BaP is independent of the solvent volume used to deliver the BaP to the skin, and therefore, independent of the exposed area.

p. D-3 (lines 20-28). This paragraph requires revision. Several pieces of information are incorrect as described below.

lines 20-23 state:

“The vehicle for benzo[a]pyrene exposure is an important factor in skin penetration. Exposure of female Sprague-Dawley rats and female rhesus monkeys topically to benzo[a]pyrene in crude oil or acetone caused approximately fourfold more extensive absorption than benzo[a]pyrene in soil (Wester et al., 1990; Yang et al., 1989).”

This statement is confusing (as to which study examined acetone compared with soil and oil alone compared with oil in soil). Also, it could imply that a factor of four is representative of BaP absorption from acetone or oil compared with soil. This is not the case. Indeed, Wester et al. 1990 also measured BaP absorption into human skin *in vitro* in which there was 17-fold more absorption into skin from acetone than from soil (same amounts of BaP were applied in acetone and in the same soil as in the *in vivo* rhesus monkey study). The Wester et al., 1990 and Yang et al., 1989 studies were quite different; to improve clarity, summaries of their study results would

be better described in two rather than one sentence. To facilitate revision, I have summarized the results of the Yang et al. 1989 and Wester et al. 1990 studies below.

In the study by Yang et al. (1989) on rats, BaP was in crude oil alone or crude oil in soil. Specifically, in both *in vitro* and *in vivo* experiments, the same mass/area of BaP (9 ng/cm²) in the same mass/area of crude oil (90 µg/cm²) was applied to skin of Sprague-Dawley rats as 100 ppm of BaP in crude oil applied directly to the skin from a solution with acetone or applied in 9 mg of soil/cm²) containing 1% crude oil. After a 96-h exposure, they observed 35.3% compared with 9.2% of the BaP absorbed into the skin (i.e., BaP measured in excreted collected during exposure and measured in skin and all other tissues at the end of exposure) from oil and soil, respectively in the *in vivo* experiments (ratio is 3.8) and 38.1% compared with 8.4%, respectively penetrated through the skin (i.e., BaP measured in the receptor chamber in the *in vitro* experiment; ratio is 4.5).

In the study from Wester et al. (1990), dermal absorption after a 24-h exposure was measured *in vivo* in rhesus monkeys and *in vitro* in excised human skin for BaP applied to skin in acetone (500 ng of BaP/cm²) or applied in soil (400 ng of BaP/cm²). Absorption into the skin determined from the two studies was different: 51% from acetone and 13.2% from soil (ratio is 3.9) *in vivo* compared with 23.79 versus 1.41 (ratio is 16.9) *in vitro*. In the *in vitro* experiments, BaP penetration through the skin was also measured from acetone (0.09% compared with 0.01% (ratio is 9). Note that a critical examination of dermal absorption from soil data concluded that the *in vivo* observation from Wester et al. (1990) may be unreliable soil (Spalt et al. 2009).

lines 25-28 state:

“Soil properties also greatly impact dermal absorption. Reduced absorption of benzo[a]pyrene occurs with increasing organic carbon content of the soil and increased soil aging (i.e., contact time between soil and chemical) (Turkall et al., 2008; Roy and Singh, 2001; Yang et al., 1989).”

The studies listed do not support the claims; in fact, in at least one case, the cited study (Turkall et al., 2008) contradicts the claim (absorption is reduced with increasing organic carbon content of the soil).

Note that the results presented in Turkall et al. 2008 were originally presented in (Abdel-Rahman et al. 2002) (unfortunately, Turkall et al. 2008 does not mention this previous publication). Turkall et al. 2008 was published as a book chapter after Abdel-Rahman et al., 2002, which was published in a journal. Therefore, it is recommended that all citations to Turkall et al. 2008 be replaced with Abdel-Rahman et al., 2002.

Abdel-Rahman et al. 2002 studied two soils that contained different amounts of organic matter (1.6% and 4.4%) that were applied to pig skin in soils that were and were not aged. Absorption of BaP in this study increased with organic carbon content, which contradicts the claim that increasing carbon content reduced absorption. Theoretically, increased organic carbon content is expected to decrease dermal absorption; however, in experiments, this has often not been the case. With respect to the effect of aging, the results from the Abdel-Rahman et al. 2002 should

be considered unreliable. It appears from the experimental description that the freshly contaminated soils were prepared by applying BaP in solvent to soil that had already been applied to the skin; thus, skin contact was to BaP in the solvent in addition to the soil (Spalt et al. 2009). I have attempted to contact the authors directly to confirm their experimental protocol, but have not received any response.

The study of penetration through human skin into the receptor fluid of the *in vitro* experiments from Roy and Singh (2001) observed no effect in soils aged 45 days and only a small decrease (~2-fold) compared with no aging. All measurements in Roy and Singh (2001) were from a single soil and did not address the effect of different organic carbon content.

The Yang et al. 1989 study did not examine effects of organic carbon content or aging.

I recently completed a thorough literature review of BaP absorption from contaminated soils, which I hope to submit for publication soon. There is no clear evidence in this literature for reduced BaP absorption with increased organic carbon content. In the only reliable study of soil aging, Roy and Singh 2001 observed a small effect in experiments on only one soil. Therefore, there are little or no empirical support for the claims of reduced BaP absorption with increasing organic carbon content of the soil and increased soil aging.

REFERENCE CITED IN REVIEW

- Abdel-Rahman, M.S., Skowronski, G.A., Turkall, R.M. (2002). Assessment of the Dermal Bioavailability of Soil-Aged Benzo(a)Pyrene. *Hum Ecol Risk Assess.* 8, 429-441.
- Booth, E.D., Loose, R.W., Watson, W.P. (1999). Effects of Solvent on DNA Adduct Formation in Skin and Lung of Cd1 Mice Exposed Cutaneously to Benzo(a)Pyrene. *Arch. Toxicol.* 73, 316-322.
- Kurihara-Bergstrom, T., Flynn, G.L., Higuchi, W.I. (1986). Physicochemical Study of Percutaneous Absorption Enhancement by Dimethyl Sulfoxide: Kinetic and Thermodynamic Determinants of Dimethyl Sulfoxide Mediated Mass Transfer of Alkanols. *J. Pharm. Sci.* 75, 479-486.
- Moody, R.P., Nadeau, B., Chu, I. (1995). In-Vivo and in-Vitro Dermal Absorption of Benzo[a]Pyrene in Rat, Guinea-Pig, Human and Tissue-Cultured Skin. *Journal Of Dermatological Science.* 9, 48-58.
- Moody, R.P., Joncas, J., Richardson, M., Chu, I. (2007a). Contaminated Soils (I): In Vitro Dermal Absorption of Benzo[a]Pyrene in Human Skin. *Journal Of Toxicology And Environmental Health-Part A-Current Issues.* 70, 1858-1865.
- Moody, R.P., Akram, M., Dickson, E., Chu, I. (2007b). In Vitro Dermal Absorption of Methyl Salicylate, Ethyl Parathion, and Malathion: First Responder Safety. *Journal Of Toxicology And Environmental Health-Part A-Current Issues.* 70, 985-999.
- Roy, T.A., Singh, R. (2001). Effect of Soil Loading and Soil Sequestration on Dermal Bioavailability of Polynuclear Aromatic Hydrocarbons. *Bull. Environ. Contam. Toxicol.* 67, 324-331.
- Roy, T.A., Yang, J.J., Krueger, A.J., Mackerer, C.R. (1992). In Vitro Percutaneous Absorption of Benzo(a)Pyrene from Crude Oil Sorbed on Soil Using Rat and Human Skin [Abstract Only]. *The Toxicol.* 12, 114.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

- Roy, T.A., Krueger, A.J., Taylor, B.B., Mauro, D.M., Goldstein, L.S. (1998). Studies Estimating the Dermal Bioavailability of Polynuclear Aromatic Hydrocarbons from Manufactured Gas Plant Tar-Contaminated Soils. *Environmental Science & Technology*. 32, 3113-3117.
- Spalt, E.W., Kissel, J.C., Shirai, J.H., Bunge, A.L. (2009). Dermal Absorption of Environmental Contaminants from Soil and Sediment: A Critical Review. *J. Expo. Sci. Environ. Epidemiol.* 19, 119-148, doi:10.1038/jes.2008.57.
- Stroo, H.E., Roy, T.A., Liban, C.B., Kreitinger, J.P. (2005). Dermal Bioavailability of Benzo[a]Pyrene on Lampblack: Implications for Risk Assessment. *Environ. Toxicol. Chem.* 24, 1568-1572.
- Stroo, H.F., Jensen, R., Loehr, R.C., Nakles, D.V., Fairbrother, A., Liban, C.B. (2000). Environmentally Acceptable Endpoints for Pahs at a Manufactured Gas Plant Site. *Environmental Science & Technology*. 34, 3831-3836.
- Turkall, R.M., Abdel-Rahman, M.S., Skowronski, G.A. (2008). Effects of Soil Matrix and Aging on the Dermal Bioavailability of Hydrocarbons and Metals in the Soil: Dermal Bioavailability of Soil Contaminants. *Proceedings of the Annual International Conference on Soils, Sediments, Water and Energy*, Article 29. 13, 403-415.
- U.S. E.P.A. (1992). Draft Report: A Cross-Species Scaling Factor for Carcinogen Risk Assessment Based on Equivalence of Mg/Kg(3/4)/Day. *Federal Register*. 57, 24151-24173, 86709.
- U.S. E.P.A. (2004). Risk Assessment Guidance for Superfund (Rags), Volume I: Human Health Evaluation Manual, (Part E: Supplemental Guidance for Dermal Risk Assessment): Final. Washington, DC.
- Wester, R.C., Maibach, H.I., Bucks, D.A., Sedik, L., Melendres, J., Liao, C., DiZio, S. (1990). Percutaneous Absorption of [¹⁴c]Ddt and [¹⁴c]Benzo[a]Pyrene from Soil. *Fundam. Appl. Toxicol.* 15, 510-516.
- Yang, J.J., Roy, T.A., Krueger, A.J., Neil, W., Mackerer, C.R. (1989a). In Vitro and in Vivo Percutaneous Absorption of Benzo[a]Pyrene from Petroleum Crude-Fortified Soil in the Rat. *Bull. Environ. Contam. Toxicol.* 43, 207-214.
- Yang, J.J., Roy, T.A., Krueger, A.J., Neil, W., Mackerer, C.R. (1989b). Percutaneous Absorption of Benzo(a)Pyrene from Soils with and without Petroleum Crude Contamination. *Petroleum Contaminated Soils*. E. J. Calabrese and P. T. Kosteki. Chelsea, MI, Lewis Publishers. 2, 399-407.

Dr. Scott Burchiel

Question 2c. Immunotoxicity (sections 1.1.3, 1.2.1). The draft assessment concludes that immunotoxicity is a potential human hazard of benzo[a]pyrene exposure. Do the available human, animal and mechanistic studies support this conclusion?

Key Points:

- The evidence for human immunotoxicity is based on complex mixture exposures; while there is no doubt that BaP and other PAHs with specific SARs can cause suppression of human HPBMC at low concentrations in vitro (10-100 nM, Davila and Burchiel, 1996), it is unclear whether these levels of exposure can be achieved from in vivo environmental inhalation exposures or ingestion of cooked foods.
- Immunotoxicity is caused by a combination of genotoxic (DNA adducts and p53 –induced cell death) and non-genotoxicity (signaling due to AhR and oxidative stress); some of these mechanisms are similar to cancer initiation and promotion, but it is unclear whether there is a threshold effect for immunotoxicity.
- Effects of BaP can vary by dose and time and sometimes leads to biphasic (U-shaped) observations of increased or decreased immune parameters, which may be mechanistically explained by differing metabolites (e.g., diol-epoxides, vs quinones) or mechanisms of action.
- Most immunotoxicity animal studies utilize mouse models (not rat) and they rely upon sensitive functional assays, such as the T-dependent antibody response (TDAR); the dose required to produce thymic atrophy are quite high in mice and rats; EPA acknowledges that thymic atrophy may not be a reliable indicator of immunotoxicity (Luster et al 1992), page 2-5, line 19.
- It is recommended that EPA establish an Immunotoxicity Guidance Document to standardize risk assessment and to identify data gaps.

Human Studies – all mixtures

- Szczeklik et al., 1994 reported decreased Ig's in serum in coke workers with mg/m³ exposures
- Zhang , 2012 studied 129 coke oven workers compared to 37 warehouse controls) for early and late apoptosis (Annexin V/PI) in HPBMC; concentrations of BaP were 10-1,600 ng/m³ in the working environment; 2,78-3.66 ng 1-OHP measured in urine
- Winker et al 1997 is an immune function and phenotype study of HPBMC comparing old and new coke facilities; results show depression of T cell activation – this study is most compelling
- Cigarette smoking - usually looking at immune suppression, but effects are complicated by the strong action of nicotine, which is immunosuppressive
- Human PBMC In vitro studies should be included in risk assessment; some studies suggest that BaP is more toxic to human HPBMC (10-100 nm) than mouse spleen cells (Davila et al, 1996); compared to 1-20 uM for mouse spleen cells White and Holsapple, 1984)

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Animal Studies

- Important structure activity relationships established early on: Dean et al (1983) showed suppression of PHA response of mouse spleen at 50 mg/kg, but not by BeP
- Consistency of findings in mice and some rat strains: Temple/White (1993) showed decreased IgM response and PFC in mouse spleen at 5, 20, 40 mg/kg and F344 rats at 10 and 40 mg/kg 14 da s,c,
- Metabolism and PK very important in BaP and other PAH immunotoxicity: General points
 - Nebert (2013) importance of balance of Cyp1A1 and Cyp1B1
 - Uno – Cyp1A1 mostly protective
 - For immunotox Cyp1B1 important in lymphoid tissues
- Immune function tests indicate that BaP is suppressive and should result in increased risk of infections and perhaps cancer; this is evidenced by Munson et al 1985 showed decreased resistance to, Strep, Herpes, and B16 melanoma by BaP but not by BeP; influenza was not affected and Listeria resistance was increased
- EPA focuses on DeJong studies in rats with toxic endpoint being thymic atrophy at 90 mg/kg; Munson Kawabata and White (1987, 1989) have shown that BaP metabolites are responsible for suppression of TDAR in mouse spleen; immune function tests are more sensitive than changes in cell viability, lymphoid organs weights, and; PAHs produce immune suppression at concentrations that are not cytotoxic

Developmental Immunotoxicity

- Since neurobehavioral endpoints are to be used for RfD calculation, I will defer to the neurotox people
- It is generally well known that developmental immunotoxicity is produced at much lower doses (10x?) than those required to produce immunotoxicity in adults; however this may not be well documented for BaP in the present literature citations used for assessment.

Other Comments:

- BaP exposures are high in woodsmoke, but there are few immunotox studies (Burchiel et al, 2005)
- We should look for evidence of increased infections in cohorts as a demonstrated health effect
- EPA should consider developing Guidelines for immunotoxicity assessment, as have been done by WHO (2012)

Dr. Anna Choi

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

Overall, the literature search strategy and process were quite comprehensive. EPA is to be commended for the first literature search for the assessment. To improve transparency, it is recommended that more specifics regarding the exclusion and inclusion criteria, and in particular, the references that were excluded and the reasons for exclusion are to be included. With regard to the study selection, it was stated that based on a secondary keyword search followed by a preliminary "manual" screen of titles or abstracts was performed by a toxicologist, and that a more detailed "manual" review of titles, abstracts, and/or papers was then conducted. Were these "manual" searches and reviews performed independently by another investigator, with search results checked and disagreements resolved? Independent searches among reviewers and the agreement and resolution of differences in studies selected is an important step in Literature Search to ensure reliability and without bias, and should be included in the assessment.

- 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?

Chen et al. (2012) was chosen as the basis for the proposed overall oral RfD. However, this study was conducted on neonate rats (post natal 6-11 days) to determine whether neurotoxic effects of postnatal BaP exposure on behavioral performance persist in juvenile and young adult stages. The draft assessment did not state whether studies on prenatal exposure assessment were considered.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Overall, the human and animal studies support that BaP is a developmental neurotoxin to human.

2c. **Immunotoxicity** (sections 1.1.3, 1.2.1). The draft assessment concludes that immunotoxicity is a potential human hazard of benzo[a]pyrene exposure. Do the available **human, animal** and **mechanistic** studies support this conclusion?

Evidence pertaining to the immune effects of BaP was based in animals (rats) (Table 1-8) studies. In particular, mainly on two studies - De Jong et al. 1999 and Kroese et al. 2001.

Some considerations on generalizability:

- *The results were based on males. What were the effects of females?*
- *The rats were 6 weeks of age. What would be the effects of mature rats? Hence, the age factor should be considered.*

Urso et al. 1988 studied the effects of BaP on spleen function. BaP exacerbates the change of the leukocyte profile during pregnancy and preferentially affects the lymphocytes. However, no follow-up was performed on the offspring of the pregnant mice.

Rats and mice appeared to have been used interchangeably on the reporting of the results. What are the differences, biologically and mechanistically, between the two that need to be kept in mind when interpreting the findings? For example, spleen effects on rats and mice (1-38), reductions in IgA levels in male rats (De Jong et al., 1999), and non-significant reductions in IgG levels in female mice (Dean et al., 1983). The evidence in animals for the effects of BaP on the developing immune system is mainly based on the studies on mice (1-40). Postnatal exposure to BaP, however, was studied on rats. The differences between the animal subjects should again be noted in concluding that BaP may alter the developing immune response to infection or vaccination.

BaP effects on the immune functions in humans are mostly based on occupational studies, and that the effects studied were mostly with PAH mixtures (except for a small number of studies such as Wu et al., 2003b which measured BaP concentrations). This should be made aware when making interpretations. BaP is often used as an indicator chemical to measure PAH exposures.

In studying the immune suppression and sensitization (1-39), a statistically significant decrease in the splenic natural killer cell activity was observed in the De Jong et al., 1999 study while no decrease was found in the Munson et al., 1985 study. The report states that the magnitude of the dose and duration of the exposure may account for the discrepancy between these two studies. It should be noted, however, that male Wistar rats were studied in the former study, and B6C3F₁ female mice were studied in the latter. The difference between the animals studied and the gender should also be noted.

The RfD for immunological effects was based on De Jong et al. (1999) where oral administration of B[a]p in male rats resulted not only in general toxicity, as indicated by the effects on body weight, but also in immunotoxicity, as indicated by the effects on bone marrow (decreased cell counts), (decrease weight in) thymus, spleen, and (decreased) lymph nodes.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Red blood cells and white blood cells were significantly decreased. Most toxic effects were only observed in the highest-dose group (90 mg/kg), but compared to the general toxicity, some parameters indicating immunotoxic effects were also affected at lower doses (10 and 30 mg/kg), including thymus weight changed and spleen B-cell.

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?

Overall, the available studies do not support noncancer toxicity in kidney and liver, although more details should be included in the conclusions.

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.

3e. **Dermal slope factor for cancer** (section 2.5). The draft assessment proposes a dermal slope factor of 0.006 per ug/day based on skin tumors in mice. 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 scaling from mice to humans? Does the method for cross-species scaling (section 2.5.4 and appendix E) reflect the appropriate scientific considerations?

2.5.1. On the human studies cited which examined skin cancer risk in relation to therapeutic coal tar (1-54 and Table D-6): limited studies were available. EPA provided a sufficient list of limitations of these studies as reasons that they do not provide an adequate basis for examining the potential association between coal tar treated patients and skin cancer. It would be more informative for reader's reference if the limitation of each study that was stated in the text would also be included in Table D-6 (with a column stating the limitation).

An additional earlier study considered for review: van Schooten et al. (1994). This was included in the WHO IPCS 1998 review.

References

van Schooten FJ, Moonen EJC, Rhijnsburger E, van Agen B, thijssen HHW, Kleinjans JCS. Dermal uptake of polycyclic aromatic hydrocarbons after hairwash with coal-tar shampoo. Lancet. 1994;1505-1506.

WHO. Selected non-heterocyclic polycyclic aromatic hydrocarbons. IPCS 1998. Geneva.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Charge question on the public comments

4. 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 by EPA.**

It appears that EPA has adequately addressed each of the comments/opinions/issues. With regard to the comment on inclusion of studies of patients therapeutically treated with coal tar, EPA's response on the limitation of the single population-based case-control study was well-stated. In addition, it should be noted that the generalizability of the results in this study is limited, as 97.7% of the subjects are Caucasians, and that male predominantly occupied the "high risk" professions.

Other comments:

1. Page 1-1, line 9: add **as** in the statement "...it is often used **as** an indicator chemical to measure exposure to PAH...."

Dr. John DiGiovanni:

1. Literature search/study selection. Is the literature search strategy well documented?

Please identify additional peer-reviewed studies that might have been missed.

In general, the literature review process is fairly well described and documented. One aspect that is difficult to assess is what information has been lost due to the exclusion of a large number of articles originally retrieved from the search criteria. In addition, a better description of the exclusion criteria might be helpful. It also seems that some more recent literature has been omitted likely due to the timeline for preparation of the document.

There are a couple of papers listed below that should be considered for citation that I did not find in the reference list:

2. Dose-response for B(a)P skin carcinogenesis in two different mouse strains Reiners, J.J. et al Carcinogenesis, 3:301-307, 1984

3. Mapping of BPDE DNA adducts in the p53 gene of NHBE cells Dessinenko, MF et al, Science 274:430-432, 1996

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 and animal studies support this conclusion?*

*2b. **Reproductive toxicity** (sections 1.1.2, 1.2.1). The draft assessment concludes that male and female reproductive effects are a human hazard of benzo[a]pyrene exposure. Do the available human and animal studies support this conclusion?*

*2c. **Immunotoxicity** (sections 1.1.3, 1.2.1). The draft assessment concludes that immunotoxicity is a potential human hazard of benzo[a]pyrene exposure. Do the available human and animal studies support this conclusion?*

*2d. **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?*

*2e. **Cancer** (sections 1.1.5, 1.2.2). The draft assessment concludes that benzo[a]pyrene is “carcinogenic to humans” by all routes of exposure. Do the available human, animal, and mechanistic studies support this conclusion?*

Collectively, the available human, animal and mechanistic data support the conclusion that B(a)P is “carcinogenic to humans”. In general, this section is well written and summarizes the available data from human epidemiology studies as well as relevant animal data and mechanistic data. The focus is on lung, bladder and skin cancers although other cancers have been linked to PAH such as B(a)P. Notably, there are no human epidemiologic studies where the exposure was to B(a)P alone. Environmental or occupational exposures to PAHs occur as mixtures of many PAHs, including B(a)P. Because there is no direct evidence in humans for B(a)P carcinogenesis this should be discussed and presented more adequately.

Note that this classification is consistent with IARC which also classifies B(a)P as carcinogenic to humans (Group 1). This classification by IARC is also based on the “strong and extensive experimental evidence for the carcinogenicity of B(a)P in many animal species, supported by the consistent and coherent mechanistic evidence from experimental and human studies that provide biological plausibility to support the overall classification of B(a)P as a human carcinogen”.

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 3×10^{-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 implicit for exposures during a critical window of development?*

*3b. **Inhalation reference concentration for effects other than cancer** (section 2.2). The draft assessment proposes an overall reference concentration of 2×10^{-6} mg/m³ based on decreased fetal survival 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 2.2.5) reflect the scientific considerations that are implicit for exposures during a critical window of development?*

*3c. **Oral slope factor for cancer** (section 2.3). The draft assessment proposes an oral slope factor of 1 per mg/kg-d based on alimentary tract tumors in mice. Is this value scientifically supported,*

giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?

Two oral carcinogenesis studies, one conducted in male and female Wistar rats and one conducted in female B6C3F1 mice, were considered for the derivation of the oral slope factor for cancer (Table 2-7). The calculated oral slope factors from these two studies (for various cancer sites) ranged from 0.04 to 1 per mg/kg-d.

The oral slope factor for cancer (1 per m/kg-day) was ultimately derived from the dose-response study of alimentary tract tumors (forestomach, esophagus, tongue and larynx combined) in female B6C3F1 mice that received oral administration of B(a)P at different doses. The rationale for selecting this slope factor is given as follows "As there are no data to support any one result as most relevant for extrapolating to humans, the most sensitive result was used to derive the oral slope factor". To base the oral slope factor for cancer on a single (albeit well designed and executed) study in only one sex seems somewhat problematic although the desire to err on the side of caution and select such a conservative value is understandable. Having data in both male and female B6C3F1 mice would be more desirable and would strengthen the determination of the oral slope using only mice. In addition, the justification for excluding the rat data and the derived slope factors from this data in the final calculation needs to be stronger.

It is clear from the oral carcinogenicity data in the studies by Kroese et al (male and female Wistar rats) and Beland and Culp (female B6C3F1 mice) as well as other data in the literature that the rodent forestomach is a very sensitive site for tumor development in both rats and mice. Because there is no human equivalent to the rodent forestomach this has been discussed in section 2.3.4 Uncertainties in the Derivation of the Oral Slope Factor and indicated in Table 2-8. It is not clear how concordance of this tumor site across mice and rats increases the relevance of the oral slope factor derivation to humans. Some additional discussion may be warranted here.

The Multistage Weibull model is used to derive the oral slope factor because it incorporates both the time at which death with tumor occurred as well as dose in the determination. While this seems to be appropriate, I think that there could be a little more detailed explanation for choosing this method over other possible methods that may have been considered. If other methods were compared what would be the oral slope factor for cancer? In the previous IRIS assessment several models were used to fit the data and a geometric mean of the estimates from 4 different models was recommended.

3d. Inhalation unit risk for cancer (section 2.4). The draft assessment proposes an inhalation unit risk of 0.5 per mg/m³ based on a combination of several types of benign and malignant tumors in hamsters. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?

3e. Dermal slope factor for cancer (section 2.5). The draft assessment proposes a *dermal slope factor of 0.006 per ug/day based on skin tumors in mice. 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 scaling from mice to humans? Does the*

method for cross-species scaling (section 2.5.4 and appendix E) reflect the appropriate scientific considerations?

The dermal slope factor (0.006 per $\mu\text{g}/\text{day}$) for skin tumors was calculated from the NIOSH study (Sivak et al, 1997) that examined B(a)P skin carcinogenesis in male C3H/HeJ mice which was then adjusted using a cross-species scaling approach similar to that used for derivation of the oral slope factor for cancer based on body weight. A number of other studies were considered but for various reasons they were excluded in the final determination.

There are significant differences in sensitivity of various mouse strains to skin carcinogenesis in addition to significant differences between species. The reliance of the derivation of the dermal slope factor on a single study using C3H/HeJ male mice leads to significant uncertainty in the calculated dermal slope factor. In addition, from the studies selected there is a suggestion of differences between male and female mice in terms of susceptibility to skin carcinogenesis by B(a)P. These uncertainties impact the derivation and accuracy of the dermal slope factor using only a single set of data from a single mouse strain and from a single sex (again in this case males). At the very least there should be some additional discussion of these two issues. Alternatively, a geometric or other type of mean could be considered by combining values from the different studies shown in Table 2.11.

As stated in the document, dermal exposure to PAH such as B(a)P in the environment likely occurs predominantly via soil contact and the available data on B(a)P carcinogenesis comes from studies where it was applied topically in a solvent. Thus, the use of a solvent may increase availability of B(a)P in the skin. Skin carcinogenesis bioassays using a soil matrix are not available to address this question, however, there are studies that have been done to examine the availability of PAH from soil and various particulates that might be considered in calculating the dermal slope factor. As with the other cancer slope factors derived in this document, B(a)P will be present in environmental samples as a mixture of many PAHs that may also impact its availability. This aspect should be considered more thoroughly in the document. Others have taken the availability of B(a)P from soil into consideration in calculating a dermal slope factor for skin cancer (Knafla et al, 2011)

In addition, there are several assumptions that are made in calculating the dermal slope factor. First, the cross-species scaling used to calculate the dermal slope factor uses the same method of allometric scaling used for the oral slope factor for cancer i.e., $3/4$ power of body weight. This method of scaling may not be appropriate for the following reasons: i) skin carcinogenesis based on dermal exposure will be proportional to the area of skin exposed as well as the dose and not necessarily proportional to body weight. Thus, scaling methods based on a unit area of skin exposed and not body weight should be considered as discussed by Knafla et al, 2011; ii) there are significant differences between mouse skin and human skin, particularly in total skin thickness as well as dermal absorption characteristics. The assumption that toxicokinetic processes in the skin will scale similarly to interspecies differences in whole body toxicokinetics is not supported by the available data. These differences need to be taken into further consideration in the cross-species scaling. Again, this is discussed in the Knafla et al, 2011 paper where an adjustment is made for these differences in skin thickness in their calculation of a dermal slope factor for skin cancer.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

3f. Age-dependent adjustment factors for cancer (section 2.6). The draft assessment proposes the application of age-dependent adjustment factors based on a determination that benzo[a]pyrene induces cancer through a mutagenic mode of action (see the mode- of-action analysis in section 1.1.5). Do the available mechanistic studies in humans and animals support a mutagenic mode of action for cancer induced by benzo[a]pyrene?

The available studies in both humans and animals support a mutagenic mode of action for B(a)P. The use of ADAFs for sensitive populations is appropriate based on a mutagenic mode of action for B(a)P.

4. Executive summary. Does the executive summary clearly and appropriately present the major conclusions of the assessment?

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

Dr. Joanne English

Charge Question 1: Literature search/study selection and Evaluation

Preliminary comments:

The process used for the literature search is clearly described. Figure LS-1 is helpful in identifying the criteria used for study selection/exclusion. In reviewing the initial literature search strategy keywords (Appendix C) it is noted that search terms for several systems (developmental, reproductive, and immunologic) were included, but no queries were made that included the term “cardio” (i.e., cardiotoxicity; cardiovascular; cardiopulmonary), “vascular,” “athero*,” etc. Given that the authors identified some evidence of cardiovascular system effects, omission of these search terms might have resulted in bias in the assessment of this endpoint. Please address.

Similarly in the literature search secondary refinement, it is noted that certain potential target organs are included in the search terms (e.g., thymus, spleen), but not others (e.g., liver, kidney). Again, it is unclear that the assessment of all potential targets identified in the hazard identification section (specifically section 1.1.4) was comprehensive and how bias was avoided. Please address if other search terms should be included.

The literature search and study selection strategy does not appear to include a review of the references in the primary literature, which is recommended as a means to surface potentially relevant articles not identified through the systematic searching and manual screening processes. Please indicate if pertinent references cited in the primary literature were reviewed, and consider including this step explicitly in the literature search and study selection strategy.

Where possible, were universal characters on a root word used to include word variations (e.g., teratog! To locate “teratogen,” “teratogenic” and “teratogenicity”)? This approach may reduce the number of search terms needed.

Additional peer-reviewed studies from the primary literature that should be considered in the assessment of noncancer health effects of benzo[a]pyrene are listed below:

Aboutabl ME, Zordoky BN , El-Kadi AO . (2009). 3-Methylcholanthrene and benzo(a)pyrene modulate cardiac cytochrome P450 gene expression and arachidonic acid metabolism in male Sprague Dawley rats . Br J Pharmacol , 158 , 1808 – 19 .

Aboutabl ME , Zordoky BN , Hammock BD , El-Kadi AO . (2011) . Inhibition of soluble epoxide hydrolase confers cardioprotection and prevents cardiac cytochrome P450 induction by benzo(a)pyrene. J Cardiovasc Pharmacol, 57, 273– 81.

Davila D, Romero D, Burchiel S . (1996). Human T cells are highly sensitive to suppression of mitogenesis by polycyclic aromatic hydrocarbons and this effect is differentially reversed by alphanaphthoflavone . Toxicol Appl Pharmacol , 139 , 333 – 41 .

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Jeng HA, Pan CH, Diawara N , Chang-Chien GP , Lin WY , Huang CT , et al . (2011) . Polycyclic aromatic hydrocarbon – induced oxidative stress and lipid peroxidation in relation to immunological alteration . *Occup Environ Med* , 68 , 653 – 8 .

Knaapen AM , Curfs DM , Pachen DM , Gottschalk RW , de Winther MP , Daemen MJ , Van Schooten FJ . (2007) . The environmental carcinogen benzo[a]pyrene induces expression of monocyte-chemoattractant protein-1 in vascular tissue: a possible role in atherogenesis . *Mutat Res* , 621 , 31 – 41 .

N ' Diaye M , Le Ferrec E , Kronenberg F , Dieplinger H , Le Vee M , Fardel O . (2009) . TNF α - and NF- κ B-dependent induction of the chemokine CCL1 in human macrophages exposed to the atherogenic lipoprotein(a) . *Life Sci* , 84 , 451 – 7 .

Oesterling E , Toborek M , Hennig B . (2008) . Benzo[a]pyrene induces intercellular adhesion molecule-1 through a caveolae and aryl hydrocarbon receptor mediated pathway . *Toxicol Appl Pharmacol* , 232 , 309 – 16.

Yang H , Zhou L , Wang Z , Roberts LJ II , Lin X , Zhao Y , Guo Z . (2009) . Overexpression of antioxidant enzymes in ApoE-deficient mice suppresses benzo(a)pyrene-accelerated atherosclerosis. *Atherosclerosis* , 207 , 51 – 8.

Wester P , Muller J , Slob W , Mohn G , Dortant P , Kroese E . (2012). Carcinogenic activity of benzo[a]pyrene in a 2 year oral study in Wistar rats. *Food Chem Toxicol*, 50, 927– 35

Charge question 2: Hazard identification (Section 1)

General comment:

It is noted in question 2, that 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 tabulated summaries for human and animal studies are organized by target organ or system effect (e.g., kidney toxicity; nervous system effects). Tabulated summaries for mechanistic studies do not appear to be included, and it is unclear how this information was integrated into the assessment of certain hazards (see comments on cardiovascular system effects, charge question 2.e. “Other types of toxicity”).

For animal studies, tabulated summaries include helpful information on study design (species, strain, sex, number per group, dose levels, route of administration and dosing regimen/duration) and study results. Additional context regarding the overall study results is often needed to interpret the findings for a specific endpoint, including available toxicokinetic information for the relevant dose range, if organ weight changes were or were not accompanied by histopathological changes; and observations that inform the general health status of animals under study.

Charge question 2c: Immunotoxicity (Sections 1.1.3, 1.2.1)

Preliminary comment:

Section 1.1.3 begins with the assertion that there are no human studies evaluating immune effects following exposure to benzo[a]pyrene alone for any route of exposure, and then discusses occupational studies. Consider including some additional human data here (Davila et al., 1996; Allan et al. 2006; Jeng et al. 2011) before animal studies are discussed.

Table 1-8 provides a clear summary of the evidence pertaining to immune effects of Benzo(a)pyrene in laboratory animals. Evidence of immunotoxicity is supported by data from multiple end-points (thymus, spleen, immunoglobulin alterations) of limited predictive capability, in combination with the mode of action analysis that suggests biological plausibility. Figure 1-5 "**Exposure-response array for immune effects following oral exposure**" nicely illustrates the NOAELs and LOAELs for repeated dose studies, showing NOAELs consistently in the 3 to 30 mg/kg-day range. This range is considerably higher than the ranges identified for developmental and reproductive toxicity endpoints. Toxicokinetic data might be helpful to inform the interpretation of the hazard data obtained in animal studies; e.g., at what dose levels is metabolic induction occurring; at what dose levels does clearance become saturated?

The authors concluded there was suggestive evidence that immunotoxicity is a potential human hazard of benzo[a]pyrene exposure. Please explain what are the criteria for the "suggestive evidence" conclusion? The preamble (Preamble 5.5) descriptor for characterizing the overall weight of evidence does not appear to be applicable to the results reviewed for immune system effects, viz.:

"Suggestive of a causal relationship: At least one high-quality epidemiologic study shows an association but other studies are inconsistent."

Please provide the relevant categories of evidential weight for causality for immune system effects, and state how the evidence for benzo(a)pyrene fulfills the criteria for "suggestive evidence" and why other levels of evidence (e.g., clear evidence or equivocal evidence) were not chosen.

Charge question 2d. Other types of toxicity (Section 1.1.4)

Preliminary comment:

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. Section 1.1.4 "Other Toxicity" begins with the statement that there is some evidence that benzo[a]pyrene can produce effects in the forestomach, liver, kidney, and cardiovascular system, as well as alter hematological parameters, but that there is less evidence for these

effects compared to organ systems described earlier in Sections 1.1.1–1.1.3 (i.e., developmental toxicity, reproductive toxicity, and immunotoxicity). Overall, EPA concluded that the available evidence does not support these noncancer effects as potential human hazards.

The potential hazards identified in the introductory paragraph; i.e., forestomach toxicity; hematological toxicity; liver toxicity; kidney toxicity; cardiovascular toxicity; as well as nervous system effects, are then discussed in the subsequent paragraphs of section 1.1.4. As noted in the response to charge question 1 - **Literature search/study selection and Evaluation**, it is unclear as to whether the search was sufficiently comprehensive to identify studies relevant to addressing the identification of all of these other hazards.

Forestomach Toxicity

The discussion of forestomach toxicity may be out of place in this section of 1.1.4 and the introductory paragraph, as there appears to be considerable evidence that the forestomach is a target of benzo(a)pyrene. The authors indicate that forestomach effects observed in rodents support a human hazard, noting that humans do not have a forestomach but do have similar squamous epithelial tissue in their oral cavity. Therefore, human relevance is not a basis for excluding the credible evidence of forestomach toxicity associated with benzo(a)pyrene exposure. As a preneoplastic (i.e., nonneoplastic) lesion, it can be logically concluded that the evidence does indeed support this noncancer toxicity as a potential human hazard. This conclusion is at odds with the overall conclusion for this section that the available evidence does not support forestomach effects as representing a potential human hazard.

In section 1.2.1 **Weight of Evidence for Effects Other than Cancer**, the authors state:

“Forestomach hyperplasia was observed following oral and inhalation exposure; however, this endpoint most likely reflects early events in the neoplastic progression of forestomach tumors following benzo[a]pyrene exposure (see Section 1.1.4), and was not considered further for dose-response analysis and the derivation of reference values.”

The authors' decision to not consider forestomach toxicity further for dose-response analysis and the derivation of reference values should not be used as a justification for excluding forestomach toxicity as a hazard credibly associated with benzo(a)pyrene exposure. Forestomach toxicity may reflect a tumor promoting key event in the tumorigenic mode of action, and thus reflect part of a combination mode of action discussed by the authors in the section “other modes of action.”

For these reasons, forestomach toxicity is credibly associated with benzo(a)pyrene exposure, so it is reasonable to identify it as such in the hazard identification section of the document. Since humans lack a forestomach, consider clarifying that in humans, such toxicity might manifest as esophageal or other gastrointestinal tract toxicity.

Hematological toxicity

The summary of hematological toxicity is well done. The data suggest that dose rate may influence blood cell parameters, but not in a reproducible fashion. Changes are minimal or statistically insignificant at all but the highest dose levels (repeated oral dosing of 90 or 100 mg/kg-day). The studies presented provide little evidence of appreciable hematotoxicity by benzo(a)pyrene. Noting the general comment to charge question 2, based on the authors' summary, I agree with the conclusion that the studies presented do not provide convincing evidence that hematological effects are a human hazard of benzo(a)pyrene exposure.

Liver toxicity

The studies described in this section reporting noncancer effects of benzo(a)pyrene to the liver can be summarized as identifying reproducible organ weight changes (all three studies) without associated histopathology in two studies. In the 3rd study, increased liver oval cell hyperplasia was reported only at the highest dose level (90 mg/kg-day) following 35-day gavage dosing (DeJong et al 1999). Clarify that histopathology evaluations of the liver were (or were not) performed by Knuckles et al. 2001. Noting the general comment to charge question 2, based on the authors' summary, I agree with the conclusion that these studies do not provide convincing evidence that noncancer liver effects are a human hazard of benzo(a)pyrene exposure. The results of Wester, et al. 2012, (not cited) should also be addressed which may provide added support for this conclusion.

Kidney toxicity

In the three studies discussed, there is no consistent finding indicative of kidney toxicity. Noting the general comment to charge question 2, based on the authors' summary, I agree with the conclusion that these studies do not provide convincing evidence that noncancer kidney effects are a human hazard of benzo(a)pyrene exposure. The results of Wester, et al. 2012, (not cited) should also be addressed which may provide added support for this conclusion.

Cardiovascular toxicity

The discussion does not convincingly lead to the conclusion that cardiovascular toxicity is not a human hazard of benzo(a)pyrene exposure. There are multiple modes of action by which chemicals may adversely impact the cardiovascular system, and it is unclear if different lines of evidence (i.e. mechanistic, animal and human) were integrated for hazard identification. Several studies showing an influence of benzo(a)pyrene on the severity and progression of atherosclerotic plaques in animal models (as cited by Oesterling et al., 2008 – not included in this section) are not addressed. Other studies to consider as part of the weight of evidence evaluation, but not cited in this section are Knappen et al (2007 and Yang et al. (2009) which address the induction of atherosclerosis by benzo(a)pyrene in rodents; and Aboutabl et al.,

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

2009 and 2011, which examine cardiac hypertrophy and cardiac biomarkers after benzo(a)pyrene exposure. The induction of inflammatory cytokines by benzo(a)pyrene (e.g.,

N'Diaye et al. 2009 – not cited; and N'Diaye et al. 2006 – cited on p 1-77) should be included as part of the weight-of-evidence discussion of cardiotoxicity. Additionally, it is unclear as to whether the designs of the animal studies reviewed were suitable to identify adverse cardiovascular effects. Although limited, the two epidemiology studies cited (Burstyn et al. 2005; Friesen et al. 2010) lend credence to possible human relevance of this endpoint.

Since cardiovascular effects were identified in rats and mice effects following gestational exposures to benzo(a)pyrene, address whether such findings should be considered as part of the weight of evidence for the cardiovascular system as a potential adult target of benzo(a)pyrene exposure.

It is unclear at this time as to whether the search was sufficiently comprehensive to identify studies relevant to addressing the identification of cardiovascular system toxicity of benzo(a)pyrene exposure (see preliminary comments to charge question 1 – literature search/study selection and evaluation). Please address the references that are missing; if they were excluded, the basis for their exclusion. If not intentionally excluded, include the missing references as part of the weight of evidence evaluation, and be explicit as to the reasoning for concluding that the available evidence either does or does not support cardiovascular system toxicity as a potential human hazard.

Nervous system effects

This paragraph briefly describes 13 articles that address nervous system effects of benzo(a)pyrene in laboratory animals and concludes with a statement that “These data are consistent with the neurobehavioral effects observed following developmental exposure and suggest that benzo(a)pyrene exposure could be neurotoxic in adults.” However, only two of these studies were identified as informing the neurotoxic potential of benzo(a)pyrene exposure in adult animals following subchronic or chronic oral exposure and included in Table 1-9. Since hazard identification does not rely only on repeated subchronic or chronic exposure scenarios alone, it is not clear why the other studies discussed in this section were not also summarized in Table 1-9; thus Grova et al. 2007; *ibid* 2008; Saunders et al. 2001, *ibid* 2002; *ibid* 2006; Liu et al. 2002 ; Maciel et al. 2014; Chen 2011; Qiu et al. 2011; Xia et al. 2011; Bouayed et al. 2012) are all left out of the table. Considering the relatively low doses in laboratory animals at which behavioral alterations were reported to be observed, the reasoning for not considering the adult nervous system as a potential human target is unclear.

Since neurobehavioral effects were identified in rats and mice effects following gestational exposures to benzo(a)pyrene, address whether such findings should be considered as part of the weight of evidence for the nervous system as a potential adult target of benzo(a)pyrene exposure.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Decrements in short term memory were reported in two studies of workers exposed occupationally to PAH mixtures containing Benzo(a)pyrene (Niu et al. 2010; Qiu et al. 2013), lending possible credence of human relevance of this endpoint.

Be explicit as to the reasoning for concluding that the available evidence either does or does not support adult nervous system effects as a potential human hazard.

Charge question 3a. Oral reference dose for effects other than cancer (section 2.1.3. Uncertainty factors)

Preliminary comments:

For all endpoints, the intraspecies uncertainty factor (UF) chosen was 10x, and it is stated that insufficient information is available to derive a quantitative estimate of variability in human susceptibility. Addressing variation within the human population, if the critical effect was in a known sensitive population, a value of less than 10 may be used. It is asserted in the document that the developing fetus is the most susceptible human subpopulation to benzo(a)pyrene; so to the extent that the rodent developing fetus is a suitable model for humans, a value less than 10x could be considered for the intraspecies UF applied in Jules et al. 2012. Thus, the default 10x factor may not warranted since the point of departure for developmental cardiovascular effects is based on exposure of a sensitive subpopulation. Since some uncertainty remains as to the variability in the susceptibility of the human developing fetus, infant and newborn to benzo(a)pyrene, a 3x UF is appropriate. A similar conclusion might be reached for the point of departure for early postnatal developmental neurobehavioral effects in Chen et al. 2012, as it is asserted under “**Susceptible Populations and Lifestages**” that the early postnatal period is also a period of heightened susceptibility to exposure to benzo(a)pyrene. However, it is as yet unclear whether the developing nervous system is more susceptible than the adult nervous system to benzo(a)pyrene exposure, based on the low dose effects in adult animals reported by Chengzhi et al. 2011 and Bouayed et al. 2009. Therefore the 10x intraspecies uncertainty factor is appropriate for the neurodevelopmental effects reported by Chen et al. 2012.

An interspecies uncertainty factor, UFA, of 3 ($10^{1/2} = 3.16$, rounded to 3) was applied, to all PODs except Chen et al. (2012), (neurobehavioral effects) because BW^{3/4} scaling is being used to extrapolate oral doses from laboratory animals to humans. Justification provided was the absence of information on whether allometric (i.e., body weight) scaling holds when extrapolating doses from neonatal animals to adult humans due to presumed toxicokinetic and/or toxicodynamic differences between lifestages. Clarify why the required extrapolation is from neonatal animals to adult humans, and not from neonatal animals to neonatal humans.

Application of subchronic to chronic UF of 1x in the case of developmental endpoints is appropriate. An UF value of 10 was applied when the POD was based on studies that were 42–90 days in duration. A value of 10x is an appropriate default for studies that are subchronic

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

(90-days) in duration, approximating 1/10 of the lifespan. For studies less than 90-days in duration, please provide justification for their use in deriving a chronic oral RfD.

Application of LOAEL to NOAEL UFs appear appropriate based on the information presented.

Selection of a database deficiency UF of 3x for all POD. The lack of a multigenerational reproduction study and lack of a neurodevelopmental toxicity study that includes exposure during gestation through lactation are identified as the data deficiencies. This is appropriate justification for the 3x UF for database sufficiency.

Dr. Michael Foster

1. Question 3b.

Inhalation reference concentration for effects other than cancer (section 2.2). The draft assessment proposed an overall ref conc of 2×10^{-6} mg/m³ based on decreased fetal survival during a critical window of development. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-analysis, calculating points of departure, and applying uncertainty factors?

Assignment: **2.2.1. Identification of studies and effects for dose-response analysis** – McIntyre, Foster, Walter.

Question 3b is focused on animal model (rodent) studies that support inhalation reference concentration for effects (developmental and reproductive toxicities) other than cancer. Reports in a female rat model were supportive of developmental (hippocampal downregulation in F-1 generation)(Wormley et al, 2004) and fetal survival (pregnant dams)(Archibong et al, 2002) effects following gestational inhalational exposures to B(a)P. A subsequent 2012 publication by AE Archibong et al, supports reproductive effects (female ovarian function) of B(a)P by the inhalation route; and although this 2012 report was not accomplished in pregnant dams, but rather non-pregnant females, the exposure methodology replicated the inhalational methodology and experimental design utilized in the earlier AE Archibong, et al, 2002 report. The AE Archibong et al, 2012 report (Endocrine disruptive actions of inhaled benzo(a)pyrene on ovarian function and fetal survival in fisher F-344 rats, *Reproductive Tox* 34:635-43) was not included as a citation and not reviewed for non-cancer effects and dose-response analysis (likely due to lateness of the report with respect to the timing of the searched publication data base prepared for the 2013 IRIS draft).

Inhalation exposure method and characterization of the B(a)P aerosol utilized by AE Archibong and colleagues (2002 report) are described with the aerosol having a tri-modal particle distribution with at least 50% of the B(a)P aerosol output in the respirable size range for the rat model utilized (Z Li, et al, PAH particles perturb prenatal processes and phenotypes: protection from deficits in object discrimination afforded by dampening of brain oxidoreductase following *in utero* exposure to inhaled benzo(a)pyrene, *Tox Sci* 125:233-247, 2012)

Additional rodent model support of developmental effects of multiple doses and inhalational exposure of B(a)P is the report by Mackenzie and Angevine (1981) and summarized in Table 1-2 (pgs. 1-5 and 1-6) using an outbred female mouse model (CD-1). This same report additionally was expanded to include reproductive effects upon F-1 generation females and subsequent viability and litter size of F-2 generation.

With respect to section **2.2.2. Methods of analysis**, in this section, on pg. 2-18, reference is made to Table 2-4 that contains summary information on female and male rat models with respect to developmental and reproductive effects (non-cancer). Appears that for some of the results summarized for male rat models and fertility outcomes, that a report by Ramesh et al,

2008 (Exp Toxicol Path 60:269-80) is not identified as a source of some of the summary information listed in Table 2-4.

Question 3b continued: does the discussion of exposure scenarios (**section 2.2.5**) reflect the scientific considerations that are **inherent** for exposures during a critical window of development?

The rodent model studies (rats and mice) appear to be highly supportive of susceptibility of F-1 generation offspring to endure developmental and reproductive effects following gestational exposures and translational to associative developmental results observed for humans (Table 1-1, pg.1.4-1.5).

2. Question 3d.

3.

Inhalation unit risk for cancer (section 2.4). The draft assessment proposes an inhalation unit risk of 0.6 per mg/m³ based on a combination of several types of benign and malignant tumors in hamsters. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure? Assignment: **2.4.1. Analysis of Carcinogenicity Data (Choice of Studies)** – Foster, Moorthy, Schlesinger

Question 3d is focused on inhalation unit risk and the assignment for 2.4.1 is to critique choice of studies that support the inhalation unit risk. Recalling the overview slide (slide #16, in the presentation by Ms. Kathleen Newhouse overviewing the Draft Assessment) during our tele-conference of March 4, 2015, a principal study was identified [Thyssen et al, 1981; experimental design: adult, male hamster model with daily (3-4.5 hr/d) life time B(a)P submicronic aerosol exposures by nose-only inhalation, over average survival durations of 60 to 96 weeks and dose response readouts of body weight, and incidence and latency of tumors with segmental distributions, i.e., URT, trachea, lung, oro-pharynx, esophagus, and forestomach]. This report was relied upon by EPA due to the merits of seemingly being the “only inhalation route cancer bioassay available” (see Executive Summary, section on Quantitative Estimate of Carcinogenic Risk from Inhalation Exposure, pg. xxxviii). Further support for reliance on the Thyssen et al, 1981, report, arises from a subsequent short communication by this same research group (see, J Pauluhn et al, 1985, Exp Path 28:31) and although limited in scope the survival results and presence of neoplastic alterations appeared replicatable with the experimental design in the hamster model for low B(a)P aerosol doses by nose-only inhalation.

Consistent with the incidence in a rodent model of URT and tracheal tumors reported by Thyssen et al, 1981, are reports and reviews from the human epidemiologic literature that demonstrate/suggest associations between exposure to PAHs related occupations and incidence of lung cancer (see Table 1-11, for summary of epidemiologic based reports of B(a)P in relation to lung cancer risk for Tier1 studies, pgs. 1-55 to 1-56) by Armstrong and Gibbs, 2009; Spinelli et al, 2006; and Xu et al, 1996, for aluminum smelter and iron-steel industry workers. For epidemiologic approaches, difficulties arise from exposure to source mixtures of PAHs and not

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

just B(a)P and confounding by habituation and/or co-exposure to cigarette smoke products. A recent 2014 review (Gibbs and Labreche, JOEM, 56: S40-S48) of epidemiologic evidence

associating increased risks of lung and bladder cancers with aluminum industry workers and occupational exposures to coal tar pitch volatiles, adds to the epidemiologic literature with a convincing association of incidence of lung cancer in workers at differing locations world-wide. Assignment: **2.5.5 Uncertainties in the Derivation of the Dermal Slope Factor** – All team members.

Derivation methods are not in my expertise and as well a significant familiarity with dermal exposure risk assessment. However background is clearly presented in the draft by EPA on the derivation of the dermal slope factor in section 2.5.5; and the Executive Summary (pg. xxxix) clearly identified the NIOSH report by Sivak and co-authors, 1997, for the data base used by EPA to dose-response analysis and extrapolation to lifetime cancer risk following dermal exposure to B(a)P. A statement in the Executive Summary (same pg. xxxix) clearly acknowledges that the dermal slope factor “has been derived for a local effect, and it is not intended to estimate systemic risk of cancer following dermal absorption of B(a)P into the systemic circulation.

4. Question 4.

Executive Summary. Does the executive summary clearly and appropriately present the major conclusions of the assessment? – McIntyres, Foster, Gennings, Li, Lichtveld, Roberts

This is a fairly broad assignment; and my comments reflect upon the section of the Executive Summary focused on inhalation exposure with effects other than cancer (pg. xxxvi), and although concise, the conveyed information in the Executive Summary for this section is appropriate and centers on fetal survival, and neurodevelopmental effects and reproductive results for both sexes citing credible animal model studies as presented in Table 1-1 on Draft pgs.1.5 to 1.7.

5. Question 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. – Bartell, Baynes, Choi, DiGiovanni, Foster, Kissell, Poirer, Portier, Roberts, Schlesinger, Stayner, Stern, Vorhees

Section G is quite extensive (~12.5 pages, text dense, with additional 1.5 pgs of an example calculation). Overall the EPA Responses seem straight forward and direct in response to the public comments. Although the public comments are grouped by section (e.g., Additional Literature, Weight of Evidence,), it would seem to be helpful and perhaps easier to scan/read through, if the public comments were identified by numbering (e.g, 1,2, 3).

Based on my own scientific background I focused on public comments in reference to Comments on the Inhalation Unit Risk (pgs. G-7 to G-9) and these comments largely dealt with

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

the Thyssen et al, 1981 report of life time exposure of B(a)P in a hamster inhalation model. Public comments related to: a) respiratory particle overload, b) discrepancies in neoplastic incidence, c) differences in the numbers of animals at risk between the EPA analysis and as

reported by Thyssen and co-authors, and d) to exposure dose variability during the course of the experimental design; for each of these comments, the responses by EPA seemed straightforward and responsive. In cases where appropriate, revisions were introduced by EPA (for example, Table D-13 of the Supplemental Information, on the incidence of benign vs malignant tumors, with respect to public comments from Arcadis and EPRI, pg. G-8).

Dr. Chris Gennings

3a. Oral reference dose for effects other than cancer (section 2.1). *The draft assessment proposes an overall reference dose of 3×10^{-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 scientific considerations that are implicit for exposures during a critical window of development?*

RESPONSE: The draft assessment does not adequately address the critical window of development. For example, the developmental toxicity studies described on page 2-2 (lines 9-10) describe oral exposures during “gestational or early postnatal development”. However, whether the studies include the critical window of development is not discussed. Similarly in the reproductive toxicity section (page 2-4) and immunotoxicity section (page 2-5), statistically significant dose-response relationships were observed in sub-chronic reproductive toxicity studies of both male and female mice; the sex of the immunotoxicity studies is not described. Seemingly, with a dose-response effect the dose range and timing of exposure were appropriate to see an effect. However, the potential impact of early-life exposure and later life effects may exacerbate these effects; this was not addressed.

The discussion of the selected studies is otherwise generally adequate; this reviewer is not aware of other important studies that should be included. The discussion around the calculation of PODs and uncertainty factors is thorough and adequate.

As the document states, “uncertainty exists due to concurrent exposure to other PAHs and other components of the mixture (such as metals).” (page 2-1, lines 29-30) However, this issue is not further addressed in considerations of exposure scenarios and the potential increase in risk due to cumulative exposure to mixtures of PAHs. This is an important omission to this section.

3c. Oral slope factor for cancer (section 2.3). *The draft assessment proposes an oral slope factor of 1 mg/kg-d based on alimentary tract tumors in mice. Is this value scientifically supported, giving consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?*

RESPONSE: The relevance of oral exposure studies in rodents is not adequately described. That is, why 2-year oral bioassay studies are relevant to human exposure should be described.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

The description of the selection of studies appropriate for dose-response analysis for calculating PODs is reasonable and seemingly adequate – although other studies may be useful. The description of uncertainty factors is appropriate. The summary tables are helpful for the reviewer.

The document states “that the oral slope factor should only be used with lifetime human exposures <0.1 mg/kg-day, because above this level, the dose-response relationship is not expected to be proportional to benzo[a]pyrene exposure.” (page 2-30, lines 23-25). The relevance of this assumption of human exposure should be further discussed, especially in consideration that the exposure is actually to mixtures of PAHs.

3d. Inhalation unit risk for cancer (section 2.4). *The draft assessment proposed an inhalation unit risk of 0.5 per mg/m³ based on a combination of several types of benign and malignant tumors in hamsters. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?*

RESPONSE: The assumptions used to derive the unit risk (that “any metabolism of benzo[a]pyrene is directly proportional to breathing rate and that the deposition rate is equal between species”; page 2-35, lines 6-8) should be evaluated. Is this a reasonable assumption?

The low-exposure extrapolation from the BMCL₁₀ in the multi-stage Weibull model used to derive the inhalation unit risk is reasonable. Appendix E provides further details. Again, the document should address how reasonable it is that lifetime human exposures will be <0.3 mg/m³ (i.e., human equivalent POD). Otherwise, the dose-response relationship is not expected to be proportional to benzo[a]pyrene exposure.

3e. Dermal slope factor for cancer (section 2.5). *The draft assessment proposes a dermal slope factor of 0.006 per ug/day based on skin tumors in mice. 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 scaling from mice to humans? Does the method for cross-species scaling (section 2.5.4 and appendix E) reflect the appropriate scientific considerations?*

RESPONSE: Based on the extensive discussion of dermal experts at the meeting, it is clear there are many issues that should be further evaluated by the EPA for developing the dermal slope factor.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

4. Executive summary. *Does the executive summary clearly and appropriately present the major conclusions of the assessment?*

RESPONSE: The primary missing part of the document is the explanation for the assessment of benzo[a]pyrene alone and not in a cumulative assessment of PAH mixtures. Presumably, benzo[a]pyrene may serve as the index chemical in a cumulative risk assessment of PAH mixtures. This should be stated. There are several places where the mixtures are described in regards to human exposure, but no clear explanation of the role of benzo[a]pyrene in the evaluation of PAH mixtures. For example, in a PAH mixture are the RfDs and RfCs of a single component adequate? This explanation should be included in the Executive Summary to set the stage for the focus on only benzo[a]pyrene in the document.

The multiple sources of exposure are described in the gray box on page 1 – however, it is not clear if there truly is a dominant source of exposure or does it really depend on human behavior.

The literature suggests that benzo[a]pyrene is an endocrine disruptor. With the description of its potential effect on birth weight, postnatal body weight and fertility it might follow. There are so many other endocrine disruptors that humans are commonly exposed to. This should be addressed. Is the effect different across sex? Should there be reference values that are sex dependent?

Several places in the ES (e.g., page xxxvii, lines 11-13) the statement is made that “confidence in the RfC is bolstered by consistent effects observed by ... similar effects observed in human populations exposed to PAH mixtures.” What does this mean? Does it imply that benzo[a]pyrene is the only active component in the PAH mixture? Does it mean the effect level is similar between the single chemical exposure in rodents to the mixture in humans, or just that there is a similar type of effect?

The document should address more directly the potential impact of early life susceptibility on later life risk of cancer and other diseases. It is mentioned on page xxxix (lines 27-29) but not adequately developed. Is there evidence of the potential degree of increased risk due to early life exposure? How relevant are the animal models used in this assessment to this issue? The use of ADAFs of 10-3-fold adjustments is stated – but is the evidence sound about this level of adjustment?

Dr. Helen Goeden

4. Literature Search/Study Selection and Evaluation.

[All members – lead discussants Goeden & Li]

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

Preliminary Response:

This section provides a concise high level, general description of the literature search strategy and study selection process conducted for benzo[a]pyrene (BaP). It should also be clear that there are additional steps beyond those portrayed in Figure LS-1. Once the chemical specific studies deemed relevant by EPA are identified additional literature to inform modes/mechanisms of action or specific areas of uncertainties may be searched for.

While Figure LS-1 provides general exclusion/inclusion criteria the reader has no efficient way to identify which specific references were excluded and why. It would be helpful if the HERO database search selection criteria could include a tag for 'not considered' along with an short explanation (e.g., duplicate). Likewise, the reader has no efficient way of understanding the utility of the retained/included references. For studies that EPA has chosen to provide study summaries in the Supplemental Information section a concise tabular summary that includes strengths/limitations and utility of the selected studies would greatly improve the clarity and transparency of the assessment.

It is noted within the document that studies with mixtures of chemicals were excluded. BaP, except in the laboratory, virtually never exists in isolation. Since this is the case risk assessments of BaP will virtually always require an assessment of BaP within a PAH mixture. At a minimum data from studies which examined both the effects of BaP alone and the effects of a PAH mixture containing BaP should be included in the current assessment as this information is type of information is essential for conducting risk characterization of BaP. EPA has undertaken an assessment of PAH mixtures and it is understandable that that effort should not be duplicated here. However, acknowledgement of what is generally known regarding synergistic, antagonistic, or additive relationships should be included in the current assessment.

This section does not describe how assessments by other national and international health organizations were identified or used within the current assessment.

Additional studies of potential interest include:

Chepelev et al. Crit Rev Toxicol 2015, 45(1):44-52. Integrating toxicogenomics into human health risk assessment: Lessons learned from the benzo[a]pyrene case study.

Moffat et al. Crit Rev Toxicol 2015, 45(1):1-43. Review Article. Comparison of toxicogenomics and traditional approaches to inform mode of action and points of departure in human health risk assessment of benzo[a]pyrene in drinking water.

Zaccaria & McClure. Int J Toxicol 2013, Jul 32(4):236-50. Using immunotoxicity information to improve cancer risk assessment for polycyclic aromatic hydrocarbon mixtures.

Zhao et al. Food Chem Toxicol 2014, Jul 69:244-251. Exposure of mice to benzo(a)pyrene impairs endometrial receptivity and reduces the number of implantation sites during early pregnancy.

Health Canada has also released a draft document: "Benzo[a]pyrene in Drinking Water" as well, which can be found at: <http://www.hc-sc.gc.ca/ewh-semt/consult/2015/bap/draft-ebauche-eng.php>

5. 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.

Preliminary Response:

Hazard identification is the process of identifying the type of hazard to human health (e.g., cancer, birth defects). Key aspects of hazard identification include identifying which health endpoints are of most concern (e.g., most sensitive - occurring at lower exposure doses than other endpoints) as well as toxicokinetics/toxicodynamics and potential MOAs as they relate to the health endpoints identified and susceptible populations (EPA 2014 Framework for Human Health Risk Assessment to Inform Decision Making).

The readability of this section would be greatly improved by the incorporation of:

- 1) An introductory paragraph which outlines the purpose of the Hazard Identification section;*
- 2) Inclusion of an overview of toxicokinetics/toxicodynamics as they relate to health endpoints and potential susceptible populations; and*
- 3) A summary paragraph which addresses the strengths and limitations of the data, including areas for which data may be unavailable (data gaps), and describes how the results of the Hazard Identification are used in the subsequent Dose Response section.*

2a. Developmental toxicity (sections 1.1.1, 1.2.1).

[Choi, McIntyres, Vorhees, Levin, Li, Poirier. Lead discussants: Levin, Vorhees]

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?

Preliminary Response:

Yes, the available human, animal, and mechanistic studies support the conclusion that BaP is a developmental hazard. The utility of the exposure-response array (Figure 1-2) could be improved by providing more context regarding dose comparisons. Currently the exposure-response arrays contain a mix of species and administration (e.g., gavage to pregnant animal, lactational exposure and direct dosing to neonatal animals) making true comparison across studies difficult. Chen et al 2012 directly dosed neonatal rats (PND5-11) whereas Bouayed et al 2009 exposed neonatal mice (PND1-14) via mother's milk and in McCallister et al 2008 fetal rats (GD14-17) were exposed in utero. This information is contained within Table 1-4 and at a minimum should be noted in Figure 1-2 to provide needed context. Calculation and presentation of Human Equivalent Doses (HEDs) should be also considered.

Toxicokinetic information regarding fetal exposures (e.g., Shendrikova and Aleksandrov, 1974. Comparative penetration of polycyclic hydrocarbons through the rat placenta into the fetus. Bull. Exp. Biol. Med., 77(2): 169–171) and lactational transfer should be included as they inform the comparative doses to developing organisms at different stages of development and exposed via different routes of administration. For example, it is likely that the neonatal animals directly dosed by Chen et al. received a higher dose than the developing organisms exposed to a comparable maternally administered dose in Bouayed et al and McCallister et al.

2b. Reproductive toxicity (sections 1.1.2, 1.2.1).

[McIntyres, Moorthy, Poirier, Walter. Lead discussants: McIntyres & Walter]

The draft assessment concludes that male and female reproductive effects are a human hazard of benzo[a]pyrene exposure. Do the available human, animal and mechanistic studies support this conclusion?

Preliminary Response:

Yes, the available human, animal, and mechanistic studies support the conclusion that BaP is a reproductive hazard. The utility of the exposure-response arrays (and possibly the evidence tables as well) could be improved by presenting Human Equivalent Dose (HED) levels. This would greatly improve the cross study comparison. Currently the exposure-response arrays contain a mix of species making true comparison across studies difficult.

2c. Immunotoxicity (sections 1.1.3, 1.2.1).

[Burchiel, Choi, English. Lead discussants: Burchiel & Choi]

The draft assessment concludes that immunotoxicity is a potential human hazard of benzo[a]pyrene exposure. Do the available human, animal and mechanistic studies support this conclusion?

Preliminary Response:

Yes, the available human, animal, and mechanistic studies support the conclusion that BaP is a potential immunotoxicity hazard.

2d. Other types of toxicity (section 1.1.4).

[Burchiel, Choi, English, Li, Ramos, Moorthy, Vorhees. Lead discussants: English & Moorthy]

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?

Preliminary Response:

The basis for arriving at this conclusion needs to be expanded for each of the health endpoints listed. The current text does not provide adequate rationale for why the evidence does not support hazard identification. Is the reason insufficient of data, inconsistent data, or sufficient data to conclude that the health endpoints is not a sensitive endpoint?

The information provided within Section 1.1.4 Forestomach Toxicity is not consistent with the conclusion drawn by EPA. There is clear evidence that BaP exposure causes forestomach hyperplasia and hyper keratosis. If it is EPA's policy that preneoplastic lesions cannot be used as the basis for deriving noncancer toxicity values or this effect is considered irrelevant to humans this should be clearly stated along with supporting rationale.

The evidence provided for hematological toxicity appears to be limited and suggests only a marginal effect on hematological parameters as the magnitude of the alterations may not be biologically significant.

The evidence provided for liver and kidney toxicity appears to be limited and suggests that while effects may be observed at higher exposure levels it does not appear to be a sensitive health endpoint.

The evidence provided for cardiovascular toxicity and adult neurotoxicity suggests potential toxicity at low dose levels, however, the data is too limited to utilize quantitatively. It is not clear why evidence pertaining to cardiovascular toxicity are not included in Table 1-9.

Relevant recently published articles include:

Gan et al. 2012. Biomed Environ Sci 25(5):549-56. Effects of benzo(a)pyrene on the contractile function of the thoracic aorta of Sprague-Dawley rats.

Jayasundara et al. 2015. Tox Sci 143(2):469-81. AHR2-Mediated Transcriptomic Responses Underlying the Synergistic Cardiac Developmental Toxicity of PAHs.

Liang et al. 2014. J Toxicol Sci 39(5):739-48. Adverse effect of sub-chronic exposure to benzo(a)pyrene and protective effect of butylated hydroxyanisole on learning and memory ability in male Sprague-Dawley rat.

Uno et al. 2014. Toxicology 316:34-42. Protective role of cytochrome P450 1A1 (CYP1A1) against benzo[a]pyrene-induced toxicity in mouse aorta.

2e. **Cancer** (sections 1.1.5, 1.2.2).

[Burchiel, DiGiovanni, Goeden, Moorthy, Poirier, Ramos, Stayner, Stern. Lead discussants: Poirier, Stayner]

The draft assessment concludes that benzo[a]pyrene is “carcinogenic to humans” by all routes of exposure. Do the available human, animal, and mechanistic studies support this conclusion?

Preliminary Response:

According to the EPA 2005 Cancer Guidelines the descriptor “carcinogenic to humans” is applied when there is strong evidence of human carcinogenicity. Combinations of the following evidence can be used to justify this classification:

Strong epidemiological evidence of an association between human exposure and either cancer or the key precursor event(s) of the mode of action but not enough for a causal association and there is extensive evidence of carcinogenicity in animals, and the mode(s) of carcinogenic action and associated key precursor events have been identified in animals, and there is strong evidence that the key precursor events that precede the cancer response in animals are anticipated to occur in humans and progress to tumors, based on available biological information.

The information provided in Section 1.2.2 and summarized in Table 1-18 does address each of the pieces of evidence necessary to identify BaP as “carcinogenic to humans”. This classification is consistent with IARC’s 2010 classification and Health Canada’s 2015 draft classification. While it is true that these assessments were prepared for different purposes, using different guidelines and methods it would be appropriate to include reference to these assessments in section 1.2.2.

6. **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).

[Bartell, Gennings, Levin, McIntyres, English, Hays, Roberts, Stern, Vorhees. Lead discussants: Stern & Bartell]

The draft assessment proposes an overall reference dose of 3×10^{-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?

Preliminary Response:

The rationale provided by EPA for not selecting Bouayed et al 2009 for dose response analysis was because the doses were higher than Chen et al. However, Chen et al used direct dosing of 0.02, 0.2 and 2 mg/kg-d to neonatal rats (PND5-11) whereas lactating mice were dosed in Bouayed et al and exposure of neonatal mice PND1-14 was via milk. It is likely that the actual doses to the developing organisms in Bouayed et al were significantly lower than the maternal doses of 2 and 10 mg/kg-d. Toxicokinetic issues such as fetal and milk transfer vs direct dosing should be discussed.

Calculation of an HED was not done for Chen et al 2012 because doses were administered directly to neonatal animals. It is true that EPA 2011 recommends that allometric scaling not be done when extrapolating doses from neonatal animals to human adults. However, EPA 2011 acknowledges that there are instances where extrapolation from the young animal to a young human exposure may be desirable. When doing such an extrapolation key developmental processes need to be matched in a species-dependent manner, because the temporal pattern of development differs across species.

It is not clear why EPA did not consider extrapolating from neonatal animals to the corresponding life stage in humans. For example (for illustrative purposes only) if body weight data for the neonatal rats from PND5-11 and humans from birth to 2 years of age is used a DAF of approximately 0.2 is calculated. Rationale for why the standard default uncertainty factor rather than extrapolating from neonatal animals to the corresponding human life stage is preferable should be added, if that is actually the case.

The rationale provided for the UF selection is reasonable. The rationale provided for application of a full 10 subchronic-to-chronic UF should be expanded. Was the available data evaluated for information regarding increased severity or additional effects or decreasing PODs with increasing duration?

3b. Inhalation reference concentration for effects other than cancer (section 2.2).

[McIntyres, Foster, Goeden, Schlesinger, Walter. Lead discussants: McIntyres & Schlesinger]

The draft assessment proposes an overall reference concentration of 2×10^{-6} mg/m³ based on decreased fetal survival 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 2.2.5) reflect the scientific considerations that are inherent for exposures during a critical window of development?

Preliminary Response:

Within the limitations of the available data the proposed overall reference concentration is scientifically supported. The exposure scenario discussion accurately reflects considerations regarding critical (and noncritical) windows of exposure.

3c. Oral slope factor for cancer (section 2.3).

[Bartell, DiGiovanni, Gennings, Portier, Roberts. Lead discussants: DiGiovanni & Portier]

The draft assessment proposes an oral slope factor of 1 per mg/kg-d based on alimentary tract tumors in mice. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?

Preliminary Response:

The proposed oral slope factor is scientifically supported. It would be relevant and appropriate to reference oral slope factors independently derived by other health organization. For example, oral slope factors recently derived by CalEPA OEHHA (2012 Public Health Goal for BaP) (1.7 per mg/kg-d) and Health Canada (2015 draft) (1.275 per mg/kg-d) are of similar magnitude.

It appears that data on mixtures from studies which evaluated both BaP alone and PAH mixtures containing BaP were excluded. BaP, except in the laboratory, virtually never exists in isolation. Since this is the case, risk assessments of BaP will virtually always require an assessment of BaP within a PAH mixture. Data from studies which examined both the effects of BaP alone and the effects of BaP within a PAH mixture (e.g., Culp et al 1998) should be included in the current assessment as this information is essential for conducting environmental risk characterization of BaP.

EPA has undertaken an assessment of PAH mixtures and it is understandable that that effort should not be duplicated here. However, acknowledgement of what is generally known regarding synergistic, antagonistic, or additive relationships should be included in the current assessment.

3d. Inhalation unit risk for cancer (section 2.4).

[Bartell, Foster, Gennings, Moorthy, Portier, Schlesinger. Lead discussants: Foster & Bartell]

The draft assessment proposes an inhalation unit risk of 0.6 per mg/m³ based on a combination of several types of benign and malignant tumors in hamsters. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?

3e. Dermal slope factor for cancer (section 2.5).

[Bartell, Baynes, Bunge, Choi, DiGiovanni, Gennings, Hays, Kissel, Portier, Roberts, Stayner. Lead discussants: Baynes & Bartell]

The draft assessment proposes a dermal slope factor of 0.006 per ug/day based on skin tumors in mice. 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 scaling from mice to humans? Does the method for cross-species scaling (section 2.5.4 and appendix E) reflect the appropriate scientific considerations?

3f. Age-dependent adjustment factors for cancer (section 2.6).

[DiGiovanni, Goeden, Poirier, Ramos, Stern. Lead discussants: Ramos & Goeden]

The draft assessment proposes the application of age-dependent adjustment factors based on a determination that benzo[a]pyrene induces cancer through a mutagenic mode of action (see the mode-of-action analysis in section 1.1.5). Do the available mechanistic studies in humans and animals support a mutagenic mode of action for cancer induced by benzo[a]pyrene?

Preliminary Response:

Yes, the available mechanistic studies support a mutagenic mode of action. Data demonstrating increased early life sensitivity to mutagenic mode of action carcinogens presented in the Supplemental Guidance (EPA 2005b) included data for BaP. Evaluations by several other health organizations have also identified mutagenicity as the primary mode of action for BaP.

7. Executive summary.

[McIntyres, Foster, Gennings, Li, Lichtveld, Roberts. Lead discussants: Li & Roberts]

Does the executive summary clearly and appropriately present the major conclusions of the assessment?

8. Charge question on the public comments

[Bartell, Baynes, Choi, DiGiovanni, Foster, Kissell, Poirier, Portier, Roberts, Schlesinger, Stayner, Stern, Vorhees. Lead discussants: Roberts & Stern]

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

Dr. Sean Hays

3c. Oral Slope Factor

- EPA has made the conclusion in the derivation of the inhalation slope factor that for route of entry contact carcinogens, BW scaling is not appropriate. I agree with this conclusion. However, in derivation of the oral slope factors, EPA has used BW scaling (BW raised to the $3/4$ power) for the oral slope factor even when the tumor sites were in the alimentary tract (route of entry). EPA should be consistent across the oral and inhalation slope factors when the tumors are at the point of contact (route of entry). It is my opinion that EPA should eliminate the BW scaling for the oral slope factors in which the tumors were in the alimentary tract. They should keep the BW scaling factor for the slope factors based on systemic tumors.

3e. Dermal Slope Factor

- I am not certain what dose metric should be used for derivation of the dermal slope factor or the appropriate means for species scaling. This seems to be an area of science that does not have good data or consensus on an approach or scaling across doses and across species. I originally thought mass/surface area (mg/cm^2) would be the most appropriate dose metric. However, it is possible that the total skin surface area contributes to the likelihood of tumors. If mg/cm^2 is multiplied by total exposure surface area, the dose metric becomes mg/day . This doesn't make sense though either since some form of species scaling should be included. The only arena I can find literature on this topic is for contact dermatitis. For this endpoint, the literature seems pretty convinced that mg/cm^2 is the most appropriate dose metric (see special issue of Regulatory Toxicology and Pharmacology Vol 52, No 1, 2008 (including Kimber et al., 2008); Felter et al., 2003). EPA's office of Pesticides has used the dose metric of mg/cm^2 as the most appropriate or regulating for contact dermatitis for exposures to chromium VI (as documented in IPCS, 2008). The question is whether the etiology of skin tumors is consistent enough with contact dermatitis from a dose metric standpoint to rely on this literature to guide the dose metric for BaP induced skin tumors. EPA has not discussed this literature and has not made a case as to whether this literature on contact dermatitis applies to BaP and skin tumors for the purposes of dose selection.
- I am also not sure that BaP painted in acetone is a reasonable surrogate for assessing the potency of PAHs in environmental media (e.g., soil).
- I agree it is reasonable to assume that the BaP applied in acetone in the mouse studies was 100% absorbed.
- It is not clear that any one of the mouse painting studies provides superior data. Therefore, I recommend EPA use averaging amongst studies.
- It is my opinion that EPA should work with interested parties (including the broader CAAC) to hold a workshop on best practices for deriving a dermal slope factor. Since

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

this is the first proposed dermal slope factor and may have very important ramifications for numerous compounds (including many that are contained in consumer products) that will be evaluated in the future, it is important that EPA get it right. A well planned workshop with interested parties would be useful for getting a broader perspective of expertise. It would be helpful to include analyses of the contact dermatitis literature and any other literature on proposed dose metrics for skin effects resulting from dermal contact.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Dr. John C. Kissel

1. Literature search/study selection. Is the literature search strategy well documented? Please identify additional peer-reviewed studies that might have been missed.

The report cites the absorption-from-soil literature uncritically. At a minimum, arguments in Spalt et al. (2009) and Kissel (2011) should be considered.

Spalt EW, JC Kissel, JH Shirai, AL Bunge. 2009. Dermal absorption of environmental contaminants from soil and sediment: A critical review. *J Expos Sci Environ Epid.* 19:119-148.

Kissel JC. 2011. The mismeasure of dermal absorption. *J Expos Sci Environ Epid.* 21(3):302-9.

5. 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.

I will confine my preliminary comments to the question of "Real World" validation of the dermal slope factor as this appears to be a key point of contention.

Comments submitted on behalf of the American Coke and Coal Chemicals Institute and others in 2013 include the following statement:

"USEPA (2013) has not performed even the most cursory real world validation of the proposed DSF to see if it makes any logical sense."

The public comment document then offers brief analyses that purport to find that EPA's analysis leads to untenable outcomes (i.e., over-prediction of skin cancer in the general population and in psoriasis patients treated with coal tar ointments). This argument was essentially repeated in the March teleconference.

EPA's response on these points is very brief and inadequate. Certainly it is reasonable to ask whether the proposed slope factor would lead to predictions that are clearly implausible. EPA declares that it cannot evaluate the public comments on the grounds that full details were not provided. However, I was able to reproduce the claims presented in the public comments (which evolved slightly between 2013 and 2015), at least in general terms, with relatively little effort.

Soil Contact. Prediction, in public comments, of high risk of skin cancer in the general populace due to soil contact is based partly on EPA arguments and partly on amendment of those arguments with additional assumptions. EPA's base case (found in Appendix G of the Supplemental Material) is described as a Central Tendency Estimate (CTE). Persons familiar with EPA practices understand that a CTE is an imprecisely defined estimate. EPA's Exposure Factors Handbook defines CTE as:

Central tendency exposure—A measure of the middle or the center of an exposure distribution. The mean is the most commonly used measure of central tendency.

In practice, since actual parameter distributions are not known, individual parameter values may be medians, means or some other statistic. When multiplied together, the results tend to be conservative, but to an unknown degree (and occasionally not at all). EPA continues to move slowly toward probabilistic approaches (a new document released by EPA's Risk Assessment Forum in 2014 states that it intends to encourage such action), but routinely reverts to deterministic approaches in cases such as this.

In the Appendix G scenario, for instance, EPA assumes all individuals contact outdoor soil or housedust indistinguishable from soil clothed in shorts and T-shirts 350 days per year for 18 years. Since the early 1990's the regulated community has complained that EPA's risk assessments suffer from "compounded conservatism," i.e., multiplication of individually conservative assumptions that in aggregate produce an outcome that represents an extreme upper tail of the population. Nevertheless, the public commenters extend that string of 350 shorts-and-T-shirts-days per year to 70 consecutive years. In concert with other assumptions unlikely to be underestimates, the public commenters produce a point estimate of risk, which they then apply to the whole population. Since that point estimate actually applies at some unspecified population fractile, it is not reasonable to assume that it applies to the whole population. EPA would be in a better position to respond if they had conducted a probabilistic assessment. This issue requires further attention from the committee.

Coal Tar Ointment Contact. The public commenters' argument is based on failure to observe elevated skin cancer in persons undergoing pharmaceutical treatment for skin disease. However, psoriasis patients are unlikely to be a relevant comparison population. Chapman et al. (1979) and Shuster et al. (1980) have found reduced AHH activity in psoriasis patients and suggested that they may be genetically less susceptible to PAH-induced carcinogenicity as a result. In addition, psoriasis patients are known to shed skin cells at greatly elevated rates (Weinstein & McCullough, 1973). Desquamation can reduce penetration beyond the stratum corneum of compounds, such as PAHs, that are lipophilic and sorb to skin cells (Reddy et al., 2000). Both mechanisms could be protective with respect to skin cancer risk due to external contact with PAHs. The finding by Roelofzen et al. (2012) of reduced 1-hydroxypyrene in urine and PAH-DNA adducts in biopsied skin in psoriasis patients in comparison to healthy volunteers following dosing with coal tar ointments is consistent with this logic. EPA discounts the pharmacological use cases due to poor quantitation of actual exposures, but fails to note that the population involved is an inherently poor surrogate for the general populace.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Dr. Ed Levin

Literature Search Strategy/Study Selection and Evaluation

Question #1

Is the literature search strategy well documented?

Yes, the authors of the review have done an excellent job describing their well-considered search strategy.

Please identify additional peer-reviewed studies that might have been missed.

Response: None identified

Hazard Identification Question #2a. Developmental Toxicity, Discussion Leaders: Levin, Vorhees

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 and animal studies support this conclusion?

Response: The evidence from the human and animal studies provides good evidence that benzo(a) pyrene exposure presents a risk for human developmental neurotoxicity. The epidemiological studies provide excellent correspondence to the more general public. Inherent in any epidemiological study there are limitations concerning the cause and effect relationship and parceling out the individual toxicants under study. The animal studies provide excellent determination of cause-and-effect relationships to individual chemicals. However the challenge of experimental animal studies is always how well do they relatedly to humans. Any study certainly has limitations but ignoring the weight of evidence the due to shortcomings of any particular study puts children at risk for neurodevelopmental disability. It is important to keep in

mind that the way we use statistical analysis is to minimize the reports of false positives (alpha errors) but most studies do not minimize the reports of false negatives (beta errors). That being the case, a particular study failing to detect a significant effect must be interpreted as a failure to detect an effect not demonstration of no effect. Inasmuch as the goal is to protect the public from toxic risks, we need to be careful not to disregard studies that find significant effects just because another study fails to find a significant effect. In fact, given that most studies are statistically powered to minimize alpha but not beta errors we should expect beta errors to occur on a regular basis.

Dose-Response Assessment Question #3a. Oral Reference Dose (RfD)

2.1.1. Identification of Studies and Effects for Dose-Response Analysis – Levin, Li, McIntyre, Vorhees

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.

Response: This section provides an excellent analysis, providing a well-considered integration of the extant literature. The Chen et al (2012) study used a within litter dosing method which could lead to cross contamination which if anything would diminish sensitivity to a BaP effect inasmuch as exposed pups could contaminate controls diminishing effects. The fact that dams may differently nurse and care for pups of different conditions could play into the effects seen. However toxicant impacts on maternal infant interactions could affect any design of this type of experiment whether they use within or between litter design. Competition between pups of different treatment conditions could exacerbate effects, but that presupposes behavioral effects of BaP to begin with. In sum I consider this study to be solid in finding BaP effects even with the handicap that controls may have had some BaP in their systems due to littermates being exposed. The use of the LSD test is not optimal. The mean and sem summary data for the water maze look robust but it is recommended that the authors be contacted for their original data so that more appropriate statistics. I recommend Dunnett's tests comparing controls to each dose group. The adverse effect of both the 0.2 and 2 mg/kg does show clear and consistent effects long term behavioral effects in the critical measure of time spent in the target quadrant (Fig. 6E).

3a. Oral reference dose for effects other than cancer (section 2.1). The draft assessment proposes an overall reference dose of 3×10^{-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

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

points of departure, and applying uncertainty factors? Does the discussion of exposure scenarios (section 2.1.5) reflect the scientific considerations that are implicit for exposures during a critical window of development?

2.1.4 – 2.1.6. All team members

Response: The authors have done a good job with the estimate of the oral reference dose for risks of non-cancer toxic consequences of BaP exposure.

Question #3b. Inhalation Reference Concentration (RfC)

3b. Inhalation reference concentration for effects other than cancer (section 2.2). The draft assessment proposes an overall reference concentration of 2×10^{-6} mg/m³ based on decreased fetal survival 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 2.2.5) reflect the scientific considerations that are implicit for exposures during a critical window of development?

2.2.4. – 2.2.8. All team members

Response: Yes, the authors of the report have done an excellent and well-considered job with this issue.

Question #3c. Oral Slope Factor

2.3.4. *Uncertainties in the Derivation of the Oral Slope Factor – All team members.*

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Response: There will always be uncertainties in the slope of the dose-effect function, particularly at the lower end of the dose range where there is a greater likelihood for variable responses. In addition non-linear dose response functions due to multiple mechanisms of effect (i.e. Ah

receptors involved in different physiological processes may have differential effects in perturbing those processes).

3c. Oral slope factor for cancer (section 2.3). The draft assessment proposes an oral slope factor of 1 per mg/kg-d based on alimentary tract tumors in mice. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?

Response: It is less than optimal to base this sort of calculation on only one line of evidence.

Question #3d. Inhalation Unit Risk

2.4.4. Uncertainties in the Derivation of the Inhalation Unit Risk – All team members

3d. Inhalation unit risk for cancer (section 2.4). The draft assessment proposes an inhalation unit risk of 0.5 per mg/m³ based on a combination of several types of benign and malignant tumors in hamsters. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?

Response: Yes I think this is appropriate way to do this calculation based on the evidence available.

Dr. Abby Li

Overall, the presentation of data in tables, including absence and presence of findings, dose levels, and 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 down side of these tables is that it's more difficult to get a sense of the consistency across findings within a study, although the supplemental information provided additional perspective. The EPA team is to be commended for the substantial effort they put forth 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 important criteria exist that may not have been fully considered in assessing the quality and utility of studies for risk assessment purposes.

9. 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 of documenting search terms used to identify studies in the main and supplementary reports. The first two 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 (third 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 that have "inadequate reporting of study methods or results" or "inadequate basis to infer exposure." However, it's not clear how EPA defines "relevant" or "inadequate." EPA could be using Sections 3.2 and 3.3, and elements of Section 4.2 and Section 6 of the Preamble, as the basis for setting a standard for 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, which are relevant for B(a)P developmental neurotoxicity studies, 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.

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.

10. 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 requires more critical evaluation of the epidemiology data, which is not the focus of this current response.

The animal literature suggests that developmental neurotoxicity is a potential hazard of B(a)P exposure at oral doses of 0.02 mg/kg/day and higher. Similar findings were reported by Chen et al. (2012; gestational exposure to rats) and Bouayed et al. (2009; postnatal lactational exposure to mice via dams) on righting reflex, negative geotaxis, and elevated plus maze. However, the experimental design has important weaknesses that limit the utility of some of the oral and inhalation developmental neurotoxicity studies for dose-response risk assessment purposes.

EPA considers the following in evaluating the quality of experimental studies (Preamble pp. xx and xxv; see also discussion above under question 1b):

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

- Documentation of study design, animals, methods, basic data, and results
- Validity of assay for its intended purpose
- Characterization of the nature and extent of impurities and contaminants
- Characterization of dose and dosing regimen
- Adequate sample sizes and statistical power (EPA guidelines require 10/sex for behavior endpoints, and statistical analysis is based on litter as the experimental unit)
- Control of variables that could influence the occurrence of effects (e.g., environmental factors, time of day, technician bias for neurobehavioral endpoints as outlined in EPA DNT, and neurotoxicity guidelines.)
- Consideration of maternal toxicity
- Relevance of animal models to humans (those that respond most like humans are preferred)
- Human route of environmental exposure
- Multiple exposure levels.

Many of these quality criteria are also included in advice to authors submitting papers for publication in the journal *Neurotoxicology and Teratology* (<http://www.elsevier.com/journals/neurotoxicology-and-teratology/0892-0362/guide-for-authors#4001>) and “[Practical considerations on the design, execution and analysis of developmental neurotoxicity studies,](#)” to be published in *Neurotoxicology and Teratology*.”

The most sensitive endpoint selected for oral risk assessment is based on Chen et al. (2012). This is a good-quality study from the perspective of executing behavioral endpoints (e.g., blind observations, randomized order of testing litters), but is 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 four dose levels. The authors clarify in the discussion section that a “**within-litter design**” was used to dose the pups, which explains 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.” The repeated rotation of dams can lead to additional stress to the pups in a manner that is not well controlled. The underlying assumption that maternal caretaking differences are randomly distributed across all litters is unverified. Poor mothers will have much greater negative impact on litters during the early pre-weaning period (PND1-10) when pups are completely dependent on the dams for warmth and nutrition. Repeated rotation of dams can exacerbate poor maternal care, and maternal care can differ if pups smell or behave differently following gavage dosing, thereby potentially favoring control pups over treated pups, as stated by Chen et al. 2012. The relative contribution and directional change based on these different factors is unknown, but it raises question about the control of environmental factors in this experiment.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

The total number of dams used, and the timing (e.g., were litters redistributed only to other dams who gave birth within 24 hr of each other?), to achieve 40 litters of 4 male and 4 female, divided into 10 litters per track, was not described in the Methods section. Presumably, all 40 litters are not born in one day, so the details on how this was achieved would be useful information when evaluating the data across ages using different "tracks" of mice.

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. From the standpoint of conducting neurobehavioral tests, this study is a weaker study than Chen et al. (2012), because (a) the sample size of five litters/dose group is not adequate for behaviors measured, and (b) 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 four pups/litter, and the authors analyzed the data with $n=20$ pups without including litter as a factor in the statistical analyses. This study is a pilot study and is inadequate for risk assessment purposes, because (a) the litter is not the experimental unit, and (b) the number of litters for each dose group is too low. 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 for anti-anxiety-like activity of chemicals, but equating increases in time in the open arm of the elevated-plus maze directly with decreased anxiety is a hypothesis that requires further testing. Nevertheless, these results are consistent with those of Chen et al. (2012).

McAllister et al. (2008) evaluated the effects of prenatal exposure (GD 14-17) on evoked potential from the primary somatic sensory (barrel) cortex following stimulation of the rat's whiskers. The litter was not the experimental unit, with a sample size of 15 pups from five to six litters for control and treated dose groups.

The Wormley et al. paper is an inhalation nose-only developmental neurotoxicity study. The restraint required in a nose-only study can induce stress in the dams, which can affect neurobehavioral effects in the offspring. For example, repeated variable prenatal stress, including restraint in a well-ventilated, cylindrical Plexiglas restrainer produces long-lasting effects on memory-induced deficits in object recognition memory, spatial reference memory using Morris water maze, conditioned fear memory, and object discrimination memory, especially working memory for objects (Markham et al. 2010, from Koenig's lab). Prenatal stress also produces effects on specific brain regions, both macroscopically and microscopically (reviewed by Charil et al. 2010).

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.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Section 1.1.4 includes neurotoxicity studies in adults. This section and Table 1-9 should include more details about the route, dose levels, and dose-response relationship.

It appears that Table 1-9 includes only the oral studies. If this is the case, EPA should indicate this in the title, because many more studies are discussed in the text. Table 1-9 also should include both positive and negative results.

The following comments focus on studies emphasized by EPA on pp. 1-49 and 2-3. Additional comments by Dr. Vorhees on the Qui et al. (2011), Xia et al. (2011), and Grova et al. (2007) studies should also be incorporated into the EPA SAB comments on "other toxicity."

It is not clear why Bouayed et al. (2012), an oral study, was not included on Table 1-9. EPA may have mistaken this as an i.p. exposure study. EPA should report the negative finding on motor activity, and indicate that there were mixed results, rather than a decreased depressive-like activity. EPA should clarify that there was no dose-response relationship (effects at 0.02 and 0.2, but not at 2 or 20 mg/kg/day), and that these effects could be acute effects, because the behavioral tests were conducted 60 minutes after gavage dosing.

EPA indicates that Bouayed et al. (2009) reported an increase in aggressive behavior and consummatory sexual behavior in mice treated with 0.02 mg/kg-day, but should indicate in the text that there were no effects at 0.2 mg/kg-day (the highest dose tested). EPA links this increase in aggressive behavior with decreased "anxiety" on the open-field test (pp. 2-3), yet the dose-response pattern is not consistent. EPA should be more cautious about interpreting these findings, because (a) the significance of four vs. two "attacks" is not clear, (b) Bouayed et al. (2009) provides no clear definition of how "attacks" were defined and distinguished from other social behaviors such as "play," and (c) the observers were not kept unaware of the treatment level.

The Grova et al. (2008) paper is an i.p. study that is not included in Table 1-9, presumably because Table 1-9 includes only oral studies. EPA relates the increased time in the open arm of the plus maze in adult animals (Grova et al. 2008) to that observed in offspring (Chen et al. 2012) (p 2-3). Yet EPA does not indicate (pp. 1-49 and 2-3) that this was a high-dose effect that occurred at 200 mg/kg (i.p.) and not at the lower doses of 0.02–20 mg/kg.

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 3×10^{-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?

EPA selected the BMDL for increased time in the open arm of the plus maze from Chen et al. (2012) as the point of departure. EPA supports this by stating that these results are consistent with similar findings following exposures of mouse pups prior to weaning (Bouayed et al. 2009a) and adult mice (Grova et al. 2008). EPA also relates this finding indirectly with increased aggression in mice (Bouayed et al. 2009b; discussed above under 2e). The dose levels and dose-response relationships should be included in this discussion of the weight of evidence.

As discussed above, there are some concerns with the Chen study. The overall quality of this study should be compared with other key endpoints/studies (e.g., cervical hyperplasia and inflammation from the developmental studies) that EPA selected for dose-response assessment.

If the Chen et al. (2012) paper is considered by EPA to be the critical study for the oral RfD, then EPA should consider escape latency from the Morris Water Maze, with the caveats described by EPA for interpretation of these results. The escape latency from the Morris Water Maze appears to be a more stable behavioral difference that was repeated over 4 days for two separate tracks (cohorts) of animals. EPA is correct that this effect is not a learning or memory effect due to difference in baseline from day 1, but is some indication of a behavioral effect (e.g. related to motor function).

5. Executive summary. Does the executive summary clearly and appropriately present the major conclusions of the assessment?

The executive summary clearly presents the major conclusions of the assessment. However, if EPA makes changes throughout the document, as recommended by EPA SAB, these changes should be reflected in the executive summary.

6. 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.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

EPA's response is partially appropriate, because biologically significant behavioral changes in different directions are appropriate as points of departure, especially if there is an overall pattern of behavioral effects. However, EPA uncritically interprets changes in behavior in terms of human emotions. EPA should objectively describe the behavior as an increase in time in the open arm of a plus maze, rather than describe this behavior as "decreased anxiety."

References not cited in EPA's report:

Charil A, Laplante DP, Vaillancourt C, King S. 2010. Prenatal stress and brain development. *Brain Research Reviews* 65:56–79.

Markham JA, Taylor AR, Taylor SB, Bell DB, Koenig JI. 2010. Characterization of the cognitive impairments induced by prenatal exposure to stress in the rat. *Frontiers in Behavioral Neuroscience* 4:1–15.

Dr. Maureen Lichtveld

Comments on executive summary

General comments:

- It is unclear who the target audience for the executive summary is. Specifically, will the same executive summary be used to inform the general public of the review findings, or will a separate literacy-competent version be developed?
- p xxxiv, lines 3-33: dark shaded text-: if the intent of that text is meant as a general overview, then the literacy level may need to be lowered. Some statements may have been oversimplified: for example, lines 14-17, the “magnitude of exposure” etc. also depends on the dose, route, and duration of course in addition to the other factors mentioned. In addition, from a community perspective there is sensitivity about always beginning this list with “lifestyle factors”, considered as a blaming strategy by some health disparate communities. While lifestyle factors certainly play an important role those could be listed as second or third.

Specific Comments

- p. xxxv line 3: *Effects other than cancer observed following oral exposure*: consider adding an example of an external measure of exposure
- p. xxxvi lines 1 and 9: *confidence in overall oral RfD*- define the qualitative term “medium”
- p. xxxviii: *quantitative estimate of carcinogenic risk from dermal exposure*- there were considerable comments from the panel members and those who presented during the public comment period of the March conference call requiring revision of this section. The details are best addressed after the upcoming in-person meeting

p. xxxix: *susceptible populations and life stages*- this section is limited in scope as presented. The *supplemental guidance for assessing early life exposure to carcinogens* has relevant information which can strengthen the current section. For example, on p. 34 of that guidance, factors influencing the analysis of susceptible life stages are outlined. A discussion of how those factors are applied in the context of B[a]P, including the derivation of the dermal slope factor, would elucidate the rationale and decision-making process regarding human health risk.

Dr. Barry McIntyre

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

The literature search approach, screening, and selection of studies for inclusion are well documented, appropriately comprehensive, and transparent. However, it is less clear as to how animal studies were assessed for quality/risk of bias. In the Literature Search Strategy, it is simply stated that "All animal studies of benzo[a]pyrene involving repeated oral, inhalation, or dermal exposure that were considered to be of acceptable quality, whether yielding positive, negative, or null results, were considered in assessing the evidence for health effects associated with chronic exposure to benzo[a]pyrene." In addition, Section 6 of the preamble identifies various factors for defining study selection for deriving the toxicity values state "credible evidence of an association". However, it is stated that "Studies with adequate power to detect effects at lower exposures levels are preferred..." This may lead one to suspect that a poorly powered study that shows an adverse outcome would be accepted. Was any consideration given to appropriately powering the study for the endpoint of interest, or using a sufficient number of animals consistent with regulatory guideline studies? Was any consideration given purity of the test material or confirmation of dose formulations? If criteria were used, were they defined a priori?

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?

P46-L 10- 15 (and others). One can't definitively state that there was an effect on fetal survival (although likely) since the authors did not stain the uteri of animals that did not litter (i.e. definitive evidence of post-implantation loss). This could be easily reworded (as appropriate) "number of litters/litter size on PND 1 was lower suggesting fetal/early perinatal loss..."

*The draft assessment includes a sufficient number of appropriately conducted animal studies that demonstrates an effect on the number of live litters/pups born (likely due to fetal death **and terata**- There are publications that indicate that B[a]P is a teratogen (see Shum et al Teratology 20(3)365 1979), growth retardation, effects on F1 fertility and fecundity when exposed in utero. These data are consistent with the limited information available from human B[a]P/PAH studies. Although there appears to be B[a]P-related changes in developmental neurological endpoints (indicative of neuro/developmental toxicity) , most of these studies utilized a small number of animals as compared to typical guideline studies (i.e. 15-20 animals/sex/group). Nevertheless, these appear to consistent with findings (e.g. cognitive ability) observed in developing humans exposed to B[a]P/PAH.*

Taken together, there is a clear and compelling relationship between B[a]P exposure and developmental toxicity in rodents. In humans, there is a compelling relationship between B[a]P/PAH exposure and fetal loss and diminished cognitive ability, and is consistent with and supported by the rodent data.

2b. Reproductive toxicity (sections 1.1.2, 1.2.1). The draft assessment concludes that male and female reproductive effects are a human hazard of benzo[a]pyrene exposure. Do the available human, animal and mechanistic studies support this conclusion?

The draft assessment includes a sufficient number of appropriately conducted animal studies that demonstrate both a functional effect on reproductive endpoints indicative of B[a]p-related reproductive toxicity, as well as evidence for potential modes of action. The rodent data demonstrates clearly that B[a]P affects fertility and fecundity. These adverse functional effects in male rodents are associated with adverse changes in the testes and sperm. The observed changes in apical reproductive endpoints (e.g. sperm motility and T) are relevant and translatable biomarkers for assessing the association of B[a]P exposure and the potential for adverse effects in humans. In human males, changes in sperm quality and fertility have been observed in individuals exposed to PAH mixtures. Although not definitive evidence of causal relationship between B[a]p exposure and reproductive toxicity in humans, these findings are consistent with those observed in laboratory animals. Studies in female rodents that may explain the functional female effects are limited, and contradictory. The Xu (2010) study was a low-powered mixture study (n=6), rather than a typical toxicity study to designed to characterize dose-response relationships and target organ toxicity. This publication has other weaknesses including the use of pentobarbital (known to affect hormone secretion), small n for low weight tissues/hormone levels. For reference, guideline toxicity studies (and studies conducted by the National Toxicology Program) typically require ~10 rats/sex. Moreover, this effect on ovarian weight was not observed by Knuckles (20 rats/group) or Kroese (10 rats/group). This being said, the study by Mackenzie and Angevine provides compelling evidence that in utero exposure (sensitive window for ovary development) to B[a]P ≥ 10 mg/kg affects the developing rodent fetal ovary, resulting infertility when the offspring are sexually mature (and in the absence of B[a]P); there is also substantial literature (e.g. Hoyer's , Mattinson's respective publications on ovarian follicle counts) the demonstrates that B[a]P has a direct effect on

the ovary and testis and provides a compelling mechanistic perspective. Findings in rodents is consistent with studies in humans examining the effects of in utero tobacco smoke and the effects on the future fertility of female offspring. Moreover, studies done by Neal et al with human tissues provides further support that the human ovary is also a target for B[a]P. The data reported by Wu et al, cannot be fully ascribed as to providing evidence that B[a]P is a human reproductive toxicant. Rather, these data are more consistent with developmental toxicity resulting in early embryonic death (which is also observed in rodents).

Taken together, there is a clear and compelling relationship between B[a]P exposure and effects on the rodent reproductive system, resulting in impaired fertility and fecundity. In humans, there is a strong relationship between B[a]P /PAH exposure and effects sperm quality and fertility, and targets the ovary, and is consistent with and supported by the rodent data.

2c. **Immunotoxicity** (sections 1.1.3, 1.2.1). The draft assessment concludes that immunotoxicity is a potential human hazard of benzo[a]pyrene exposure. Do the available human, animal and mechanistic studies support this conclusion?

2d. **Cancer** (sections 1.1.5, 1.2.2). The draft assessment concludes that benzo[a]pyrene is "carcinogenic to humans" by all routes of exposure. Do the available human, animal, and mechanistic studies support this conclusion?

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?

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 3×10^{-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?

In principal, the selection of an overall reference dose based on developmental toxicity during a critical window of development is scientifically supported, but the selection of studies upon which it is based warrants further panel discussion. In the study by Chen

(2012), one could argue that there are potential study design/conduct weaknesses that may decrease the confidence in the study findings. These include: rotating the pups amongst dams every few days, relatively small sample size (as compared to guideline studies), and potentially inappropriate statistical analyses. Specifically, rotation of pups amongst dams likely induces both pup and dam stress (very common for dams to reject their fostered young). Moreover, it does not "distribute" the maternal caretaking differences across the groups; it actually results in the loss of the ability to account or test for maternal/litter effects. This study utilized 10 animals/gender/group; the authors did test for gender effects, and polled if not significant (for an n of 20). For righting reflex (which exhibited an effect at the lowest dose) the effect of treatment x gender exhibited a p value of 0.10, whereas gender alone exhibited a p value of 0.06. Taken together, these appear to approach statistical significance, and may have attained the $p < 0.05$ if more animals were used, and resulting data may be confounded by gender.

The study by Xu (2010) was a low-powered mixture study ($n=6$), rather than a typical toxicity study designed to characterize dose-response relationships and target organ toxicity. This publication has other weaknesses including the use of pentobarbital (known to affect hormone secretion), small n for low weight tissue/hormone levels. Moreover, this effect on ovarian weight was not observed by Knuckles (20 rats/group) or Kroese (10 rats/group) at similar and higher dose levels. Therefore, the selection of this study for further dose-response analysis may not be appropriate. Obviously, study selection will impact (to some degree) the subsequent presentation and collective assessment. Conceptually, the assessment of candidate values, UFs and PODs is logical and appropriate. However, The EPA should further justify the application of an UFd of 3 (currently stated because a multi-gen or OECD 443 was not available). The current data base could be considered sufficient as multigenerational studies were conducted and demonstrated adverse outcomes that are supported by mode of action studies. With the advent of the extended one generation design (OECD 443- which is a considered a replacement for the multi-gen), F1 animals, which have been continually dosed, are only assessed for reproductive effects if triggered (Parental generation is only required to be dosed for 2-weeks prior to mating). Therefore, it is questionable that a standard OECD 443 will provide any additionally useful reproductive information.

3b. Inhalation reference concentration for effects other than cancer (section 2.2). The draft assessment proposes an overall reference concentration of 2×10^{-6} mg/m³ based on decreased fetal survival 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 2.2.5) reflect the scientific considerations that are inherent for exposures during a critical window of development?

In principal, the selection of an overall reference concentration based on adverse effects during a critical window of development is scientifically supported, but the study selected appears to have deficiencies that warrant further panel discussion. The selected study by

Archibong (2002) exhibits technical weakness that may impact overall study consideration. Blood samples were collected from the orbital plexus (a highly stressful technique); stress is known to elevate PRL levels (potential relationship between B[a]P and PRL is suspect), and based on the hormone data it appears that each dose (and corresponding control) was run in series (each dose group having its own control with discrepancies in control responses). These weaknesses aside, the apparent effects of B[a]P on fetal survival are compelling, and consistent with that observed in other studies. Obviously, study selection will impact (to some degree) the subsequent presentation and collective assessment. EPA should provide additional justification for study selection/deselection, use of BMD modelling, and application of UFs.

3c. Oral slope factor for cancer (section 2.3). The draft assessment proposes an oral slope factor of 1 per mg/kg-d based on alimentary tract tumors in mice. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?

3d. Inhalation unit risk for cancer (section 2.4). The draft assessment proposes an inhalation unit risk of 0.6 per mg/m³ based on a combination of several types of benign and malignant tumors in hamsters. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?

3e. Dermal slope factor for cancer (section 2.5). The draft assessment proposes a dermal slope factor of 0.006 per ug/day based on skin tumors in mice. 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 scaling from mice to humans? Does the method for cross-species scaling (section 2.5.4 and appendix E) reflect the appropriate scientific considerations?

3f. Age-dependent adjustment factors for cancer (section 2.6). The draft assessment proposes the application of age-dependent adjustment factors based on a determination that benzo[a]pyrene induces cancer through a mutagenic mode of action (see the mode-of-action analysis in section 1.1.5). Do the available mechanistic studies in humans and animals support a mutagenic mode of action for cancer induced by benzo[a]pyrene?

4. Executive summary. Does the executive summary clearly and appropriately present the major conclusions of the assessment?

The executive summary clearly presents the current major conclusions of the assessment (based on the selected studies that were used for calculation of the RfD, RfC, and cancer slope factors). The Summary also addresses the Key Issues and provides the context in which these were addressed.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

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

Dr. Bhagavatula Moorthy

Charge questions on the draft Toxicological Review

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

Comments: The overall strategy for literature search/study is appropriate. However, some references could be included in regard to the effect of maternal exposure to benzo[a]pyrene (BP) on fetal development. Recent epidemiological studies suggest an association between dietary BP intake and lower birth weight in children (Duarte-Salles et al., *Environment international* 60C, 217-223, 2013; Duarte-Salles et al., *Environment international* 45, 1-8, 2012; *Public health nutrition* 13, 2034-2043, 2010). These references could be included. Also, there is little emphasis on the effects of benzo[a]pyrene (BP) on non-cancer pulmonary toxicity. Our group recently published a paper in which we showed that maternal exposure of mice to BP leads to increased susceptibility of newborn mice to hyperoxic lung injury and chronic lung disease (CLD) (Couroucli et al., *Tox. Lett.*, 230: 322-332, 2014). Supplemental oxygen therapy is frequently encountered in premature infants and very low birth weight infants, and hyperoxia contributes to the development of bronchopulmonary dysplasia (BPD), also known as CLD, in these infants. Maternal smoking is one of the risk factors for preterm birth and for the development of bronchopulmonary dysplasia (BPD). Thus, I believe there should be some description on the effect of BP on pulmonary toxicity in infants as well as adults.

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?

Comments: Yes, this is true. I will also add that additional emphasis could be given on the effect of BP on lung development because maternal exposure to BP through cigarette smoke or diet (e.g., charcoal broiled meats) could lead to abnormal lung development in the babies born to these mothers.

2b. Reproductive toxicity (sections 1.1.2, 1.2.1). The draft assessment concludes that male and female reproductive effects are a human hazard of benzo[a]pyrene exposure. Do the available human, animal and mechanistic studies support this conclusion?

Comments: Yes, this is true. In a recent study ([PLoS One](#). 2014 Jan 29;9(1):e87439. doi: 10.1371/journal.pone.0087439), male transgenic lacI mice at different ages (7, 25 and 60 days old) were treated with BP at different doses (0, 50, 200 or 300 mg/kg body weight). Mutant frequency was then determined in a meiotic cell type (pachytene spermatocyte), a post-meiotic cell type (round spermatid) and epididymal spermatozoa after at least one cycle of spermatogenesis. The results show that (i) mice treated with BP at 7 or 25 days old, both being pre-adult ages, had significantly increased mutant frequencies in all spermatogenic cell types tested when they were 60 days old; (ii) spermatogenic cells from mice treated before puberty were more susceptible to BP-associated mutagenesis compared to adult mice; and (iii) unexpectedly, epididymal spermatozoa had the highest mutant frequency among the spermatogenic cell types tested. These data support the hypothesis that pre-adult exposure to BP increases the male germline mutant frequency in young adulthood. The data also suggest that exposure to environmental genotoxins at different life phases (e.g., pre-adult and adult) can have differential effects on reproductive health. This information could be included in the draft. In regard to females, Finuadi et al. ([Hum Reprod](#). 2014 Mar;29(3):548-54) showed that *in vivo* exposure to benzo(a)pyrene induces significant DNA damage in mouse oocytes and cumulus cells.

2c. Immunotoxicity (sections 1.1.3, 1.2.1). The draft assessment concludes that immunotoxicity is a potential human hazard of benzo[a]pyrene exposure. Do the available human, animal and mechanistic studies support this conclusion?

2d. Cancer (sections 1.1.5, 1.2.2). The draft assessment concludes that benzo[a]pyrene is "carcinogenic to humans" by all routes of exposure. Do the available human, animal, and mechanistic studies support this conclusion?

Comments: Yes, the data do suggest that BP is a human carcinogen. The section is well organized and well written. While BP is understood to be a carcinogen by all routes of exposure, it needs to be emphasized that BP is not a liver carcinogen, and mainly causes cancer in organs such as lung, breast, skin and the route of exposure determines where the primary site of cancer formation is likely to occur. For example, dermal exposure leads to skin cancer, while inhalation exposure could lead to lung cancer, and chewing tobacco could lead to oral cancer. Also, while liver cytochrome P4501A/1B enzymes play a major role in the bioactivation of BP to form metabolites that can bind DNA, leading to DNA adducts, these hepatic adducts could not cause cancer in the liver. Actually some of the metabolites could be transported to lung or there is local metabolism in lung leading to adducts which cause tumorigenesis in the lung. We have a review article that is in press in *Toxicological Sciences* that in part discusses the latest aspects of PAH metabolism and lung cancer (Moorthy et al., Polycyclic Aromatic Hydrocarbons (PAHs): From Metabolism to Lung Cancer, *Toxicological Sciences*, in press, 2015).

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?

Comments: Yes, please see my comments in response to charge question 1.

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 3×10^{-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?

Comments: Yes, this has been well presented. However, please see my comments in response to charge question 1 for the relationship between maternal BP and chronic lung disease in newborn mice.

3b. **Inhalation reference concentration for effects other than cancer** (section 2.2). The draft assessment proposes an overall reference concentration of 2×10^{-6} mg/m³ based on decreased fetal survival 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 2.2.5) reflect the scientific considerations that are inherent for exposures during a critical window of development?

3c. **Oral slope factor for cancer** (section 2.3). The draft assessment proposes an oral slope factor of 1 per mg/kg-d based on alimentary tract tumors in mice. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?

3d. **Inhalation unit risk for cancer** (section 2.4). The draft assessment proposes an inhalation unit risk of 0.6 per mg/m³ based on a combination of several types of benign and malignant tumors in hamsters. Is this value scientifically supported, giving due consideration to the

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?

Comments: While this may be appropriate, I believe additional studies in different species (e.g., rats, mice) must be done for extrapolation to humans.

3e. **Dermal slope factor for cancer** (section 2.5). The draft assessment proposes a dermal slope factor of 0.006 per ug/day based on skin tumors in mice. 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 scaling from mice to humans? Does the method for cross-species scaling (section 2.5.4 and appendix E) reflect the appropriate scientific considerations?

3f. **Age-dependent adjustment factors for cancer** (section 2.6). The draft assessment proposes the application of age-dependent adjustment factors based on a determination that benzo[a]pyrene induces cancer through a mutagenic mode of action (see the mode-of-action analysis in section 1.1.5). Do the available mechanistic studies in humans and animals support a mutagenic mode of action for cancer induced by benzo[a]pyrene?

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

Miriam C. Poirier

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

Response - The approach to the literature search used for this evaluation has been described in the Preamble on page xvii, Section 3. The major standard databases (PubMed, ToxNet, NLM, Web of Science), as well as the EPA's HERO database were queried for items including benzo[a]pyrene, PAHs, and other relevant terms, with interest in animal data, human epidemiological data and other mechanistic and toxicokinetic studies. The original searches turned up 21,000 references, and those were pared down to about 700 references, which are currently cited in the Toxicological Review and fall into 7 different categories. All of these references are currently available in the HERO database.

Given the task at hand I believe that the EPA has done a fine job with this. However, given the incompleteness of modern databases and the variety of terms required to search this very complex topic, it is not surprising that there were occasional relevant papers missed. In addition there may be papers that have been left out intentionally due to evaluation criteria that might not be clear to those of us reviewing the document. Having a panel of experts review the literature chosen is an important aspect of this review, given the abovementioned difficulties in finding all the relevant papers.

Conclusions - Several papers relevant to this document were not found in the HERO database. They have been included in the discussion of the charge questions below.

Charge Question 2d. Cancer (sections 1.1.5 and 1.2.2) – “The draft assessment concludes that benzo[a]pyrene is ‘carcinogenic to humans’ by all routes of exposure. Do the available human, animal and mechanistic studies support this conclusion?”

Response – To answer this it is necessary to invoke the EPA guidelines for whether or not a compound is considered a human carcinogen. These are evaluated below with respect to the evidence presented in the “Toxicological Review of Benzo[a]pyrene”.

The compound in question is “Carcinogenic to Humans” when there is convincing epidemiologic evidence of a causal association between human exposure and cancer.

The epidemiologic data, presented here in the “Toxicological Review” (p.1-83 and 1-84), summarize a large number of studies focused on lung, bladder and non-melanoma skin cancer, and strongly support the carcinogenicity of polycyclic aromatic hydrocarbon (PAH) exposures in humans. However, as this report states, in the arena of human exposure, it is not possible to separate benzo[a]pyrene from other carcinogenic PAHs. Therefore, from the epidemiologic studies there is no direct evidence that benzo[a]pyrene by itself (alone) is carcinogenic to humans. However, because there is the assumption that benzo[a]pyrene is a likely component of all the PAH mixtures that humans are exposed to, benzo[a]pyrene alone can be considered a human carcinogen.

The Toxicological Review document focused on lung, bladder and skin cancers, but there are other organs for which PAHs are carcinogenic (see Supplemental information, p. D-28 to D-33). For example, colon cancer risk. There is strong evidence for an association between PAH-exposure in heavily char-broiled meat (Rothman et al., HERO ID 84099) and colon adenoma risk (Sinha, R. et al., HERO ID 1007703). In addition there are strong associations between PAH-DNA adduct formation, cooked meat ingestion and colon adenoma risk in the same population (Gunter et al., HERO ID 1011897).

The compound in question can be considered “Carcinogenic to Humans” when there is a lesser weight of epidemiological evidence but when all of the following conditions are met:

a) strong evidence of an association between human exposure and either cancer or the key precursor events of the agent’s mode of action but not enough for a causal association

The Toxicological Review rightly concludes that benzo[a]pyrene is metabolized, damages DNA, and is carcinogenic through consequent mutagenic mechanisms, which lead to tumor formation. The document fails to point out, however, that of these steps (metabolism, DNA damage, mutagenesis) the only one that can be demonstrated specific to PAH exposure is the formation of PAH-DNA damage. Because many classes of carcinogen (in addition to PAHs) induce the formation of GC→TA and AT→TA transversions, not a single human mutation can be unequivocally traced back to a PAH exposure. Therefore, despite the indirect evidence presented in the Toxicological Review on p.1-84, the evaluation of mutation spectra alone cannot indicate the cause of those mutations.

In humans the presence of PAH-DNA adducts is a critical step in the continuum between exposure and tumor induction, however specific evidence for benzo[a]pyrene-induced DNA adduct formation is more difficult to find in humans. In an excellent review (Boysen and Hecht, *“Analysis of DNA and protein adducts of benzo[a]pyrene in human tissues using structure-specific methods”* Mutation Research 543:17-30, 2003 – which is not in the HERO database), the authors document the use of structure-specific methods to quantify benzo[a]pyrene-induced DNA adducts in human tissues. They reported that in 39% of 705 human samples it was possible to detect the presence of the major stable DNA adduct associated with benzo[a]pyrene exposure and carcinogenesis, the r7, t8, t9-trihydroxy-c-10-(N²deoxyguanosyl)-7, 8, 9, 10-tetrahydro-benzo[a]pyrene (BPdG). In conclusion, BPdG formation in human tissues provides a direct link between PAH exposures and mutations considered likely to be associated with tumor risk.

Additional comments:

The Supplemental information summarizes six human studies (Table D-33), which evaluated benzo[a]pyrene-induced DNA adducts in humans. This is a small fraction of the available studies that employ chemical class-specific methods to measure PAH-DNA and BPdG adduct formation in human tissues. It is possible that some epidemiological studies have been omitted by the EPA because for lack personal monitoring data for each individual. However, one could argue that individuals in a workplace that is known to be polluted should not require personal monitors because the presence of high levels of urinary 8-hydroxy-pyrene, or of the BPdG adduct in any organ, are also good indicators of exposure. I would urge the EPA to reconsider the requirement for individual monitoring data in epidemiological studies, although I would agree that dosing in animal studies should be exact.

There are a series of human epidemiological studies, involving small (>100) cohorts of individuals, where subjects have been stratified into quartiles or quintiles for their PAH-DNA adduct level (using chemical class-specific methods). These studies have found significant increases in cancer risk in individuals with the highest PAH-DNA adduct levels. This data would make a useful table in the Supplemental information (see: Kyrtopoulos, S.A., Toxicology Letters 162:3-15, 2006 [not in HERO]; and Poirier, M.C., HERO ID 2558407).

Critical to our understanding of the published values for human BPdG and PAH-DNA adducts, is knowledge of what is actually being measured by a specific assay. The gold standard is determination by structure-specific methods (mentioned above). Other assays can have compound-class specificity. For example, the various antibody-based methods (ELISA and immunohistochemistry) employ monoclonal or polyclonal antibodies (termed BPDE-DNA antisera) raised against benzo[a]pyrene-modified DNA, which cross-react with a family of carcinogenic PAHs bound to DNA. We use the term “PAH-DNA adducts” for measurements of human tissue DNA using these antisera because multiple carcinogenic PAH-DNA adducts are measured. Other assays are not PAH specific. For example ³²P-postlabelling, which detects adducts of many different chemical classes, is not at all specific for BPdG when using human samples. Choice of an assay will vastly impact the validity, reliability and conclusions obtained from a particular study. The Toxicological Review has no consistent discrimination between the various methods used for human PAH-DNA and/or BPdG analysis. This may be due to the lack

of a consistent chemically-correct nomenclature on the part of the authors themselves. The Toxicological Review and Supplemental information could be made much more clear if a table could be added to describe the characteristics and nomenclature of the methodologies in question.

b) extensive evidence of carcinogenicity in animals

The document, on pages 1-62 to 1-69, and the summary on pages 1-85 and 1-86, provide a thorough documentation of many different studies all showing unequivocally that benzo[a]pyrene is a carcinogen in rodent models.

c) the mode(s) of carcinogenic action and associated key precursor events have been identified in animals

As demonstrated clearly in the Toxicological Review (and summarized in Supplemental Information Table D-33, p.D-98), in animal models exposed to benzo[a]pyrene there is extensive evidence of the formation of dose-related benzo[a]pyrene-induced DNA adducts, and other types of dose-related genotoxic events including: germline mutations, somatic mutations, micronuclei, sister-chromatid exchanges, chromosomal aberration, DNA strand breaks, and unscheduled DNA synthesis.

d) there is strong evidence that the key precursor events that precede the cancer response in animals are anticipated to occur in humans and progress to tumors, based on available biological information

Mechanistic evidence in human tissues supports the diol-epoxide pathway as leading to PAH-DNA and BPdG adduct formation, which are key events linking PAH exposures with cancer-associated mutations GC→TA transversion mutations. However, because benzo[a]pyrene is a complete carcinogen, with initiating and promoting activities, other end points, which also occur in a dose-related fashion in animals, may play a role in tumor induction. These are likely to include DNA damage occurring through the radical cation and the *o*-quinone pathways, chromosomal aberrations, sister chromatid exchanges and micronucleus formation (Supplemental Information page D-26). In addition, mutations in glycophorin A and *HPRT* are associated with PAH exposure in humans.

Conclusions – The first step in the EPA analysis of whether or not a compound is a human carcinogen states “***The compound in question is ‘Carcinogenic to Humans’ when there is convincing epidemiologic evidence of a causal association between human exposure and cancer***”. Whereas the available epidemiological data do show that PAHs are carcinogenic in humans, there is no data on human exposure to benzo[a]pyrene alone. The strong possibility that all PAH mixtures contain benzo[a]pyrene provides a likelihood that this is the case. For the second step of the EPA requirements, “***The compound in question can be considered ‘Carcinogenic to Humans’ when there is a lesser weight of epidemiological evidence but when all of the following conditions are met....***”, the data show that all four of the required conditions are met. Therefore, based on tumor studies in humans and animal models, and on mechanisms

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

of action determined in both species, strong evidence of key precursor events related to benzo[*a*]pyrene exposure and found in humans indicates that benzo[*a*]pyrene can be considered a human carcinogen.

Charge question 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?

Response – As the Toxicological Review states on p. 1-1, section 1.1.1, there is evidence that in human pregnancies where the mother is exposed to ambient or dietary PAHs, PAH-DNA damage as an indicator of exposure is accompanied by: a reduction in fetal size, an increase in *in utero* fetal loss, a reduction in the size of head circumference (related to cognitive function), developmental delay in motor skills, increased anxiety/depression, and increased attention problems. With the exception of increased tendency to abort, and reduction in fetal size, all of the other end points are considered manifestations of neurotoxicity. Clearly these toxicities may have significant impact on the lives and future success of these children.

Many rodent studies support the human observations, though in a few areas the results diverge. In the rodent studies benzo[a]pyrene treatments during gestation were either by gavage or inhalation. In parallel to the human studies, there were significant increases in fetal loss, along with decreases in fetal weight and survival. In addition, examination of fertility in the offspring revealed changes in development of male and female reproductive organs, and decreased overall fertility which included fewer and smaller litters, compared to unexposed controls. Physical examination revealed organ weight decreases and abnormalities in testes and ovaries. Additionally, cardiovascular and neurological defects were observed, including increased blood pressure, altered learning and memory behaviors, and impaired neuromuscular and sensorimotor development. However in the rodents, unlike in the children, there was decreased anxiety-like behavior.

With the exception of anxiety-like behavior, all of the end points found in children have been reproduced in rodent models. Additional end points found in the rodents, for example the reproductive integrity and fertility-related issues, have yet to be documented in children, but are indicators of potential long-term consequences of PAH exposures in children.

Conclusions - This part of the Toxicological Review document is comprehensive and well-written, and I have no suggestions for additional studies or references. The data support the conclusions that developmental toxicity in both males and females are likely outcomes from transplacental benzo[a]pyrene exposure.

Charge question 2b. Reproductive toxicity (sections 1.1.2, 1.2.1). The draft assessment concludes that male and female reproductive effects are a human hazard of benzo[a]pyrene exposure. Do the available human, animal and mechanistic studies support this conclusion?

Response – As the Toxicological Review states on p.1-22 in section 1.1.2, in humans there is evidence that environmental and occupational PAH exposures affect both male and female fertility. In two studies PAH-exposed males were shown to have low fertility, with their embryos having abnormally low implantation rates. Workers occupationally exposed to PAHs were more likely to have oligospermia and morphologically-abnormal sperm, than controls. In females, women who smoke are more likely to have ovulatory disorders and higher spontaneous abortion rates than women who do not. In addition, smoking during peri-menopause has been shown to accelerate the rate of menopause.

In rodent models, where benzo[a]pyrene exposure was largely by gavage or inhalation, the evidence is much more extensive. Changes in males were found in several studies, and these included decreases in sperm counts, decreases in sperm motility and altered sperm morphology. In addition, there were decreases in testicular weight, decreased epididymal tubule diameter, and decreases in testosterone levels. In benzo[a]pyrene exposed females, decreased fertility and fecundity, decreased ovary weight, decreases in the number of follicles, and a reduction in follicle stimulating hormone (FSH) were observed. In pregnant mice, benzo[a]pyrene exposure exposure has been shown to produce losses in progesterone, estradiol and prolactin. Also noted in females were altered estrus cyclicity and increased cervical epithelial inflammation.

Taken together the rodent studies support the observations of reduced fertility in human smokers and workers exposed to high levels of benzo[a]pyrene in PAH mixtures. The rodent studies add critical mechanistic insights that could not be obtained from the available human studies alone.

Conclusions - This part of the Toxicological Review document is comprehensive and well-written, and I have no suggestions for additional studies or references. The data support the conclusions that male and female reproductive effects are a human hazard of benzo[a]pyrene exposure.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Charge question 3f. age-dependent adjustment factors for cancer (section 2.6). The draft assessment proposes the application of age-dependent adjustment factors based on a determination that benzo[a]pyrene induces cancer through a mutagenic mode of action (see the mode-of-action analysis in section 1.1.5). Do the available mechanistic studies in humans and animals support a mutagenic mode of action for cancer induced by benzo[a]pyrene?

Response – The document EPA/630/R-03/003F “Supplemental Guidance for Assessing Susceptibility from Early-Life Exposures to Carcinogens” lays out the rational approach of the EPA, to adjustment of tumor risk for exposures at different ages, for carcinogens with a mutagenic mode of action. The age-related adjustments are based on many animal experiments where exposures occurred at different ages, and tumor incidences were evaluated.

Conclusions - Having read this document, and section 2.6 in the Toxicological Review of Benzo[a]pyrene, I agree that use of the proposed age-dependent adjustment factors (ADAFs) in evaluation of exposures in human infants and adolescents is science-based and reasonable. In addition the supporting references are complete, and there is nothing else I would add.

Dr. Kenneth Portier

Question #1: Literature Search Strategy/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. Please comment on whether the literature search approach, screening, evaluation, and selection of studies for inclusion in the assessment are clearly described and supported. 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. Please 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

Preliminary Comments:

- *Figure LS-1 references conditions that led to the exclusion of ~600 manuscripts in the manual screening of the ~1000 manuscripts considered for study inclusion. These criteria are not described or otherwise expanded upon in the Literature Search Strategy section of the BaP Tox Review document (the Report) or in the BaP Tox Supplemental Information (the Supplement). In particular, phrases such as “Inadequate basis” and “Inadequate reporting” sound subjective but it is assumed that some additional criteria were used to make this judgement (e.g. used no controls, inappropriate route of exposure, very low or no power to detect effects, etc.) While these issues may be discussed in A Review of the Reference Dose and Reference Concentration Processes (U.S. EPA, 2002) and Methods for Derivation of Inhalation Reference Concentrations and Application of Inhaled Dosimetry (U.S. EPA, 1994)., at a minimum, the BaP Tox Supplemental Information should be appended to include additional discussion of this issue. Without this additional information, the Study Selection cannot be duplicated. I agree that much of the detailed rationale for selected studies is provided in the Dose Response subsections of each section.*
- *No assessment is provided on the extent of information lost by not including animal in vivo and in vitro studies designed to identify potential therapeutic agents that would prevent the carcinogenicity or genotoxicity of benzo[a]pyrene. It is to be expected such studies might provide additional information on mode of action of benzo[a]pyrene. It may be that the extensive discussion in Appendix D covers everything that might be relevant from these therapeutic agent animal studies, but if this is the case EPA should so state.*

Question #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?

2b. Reproductive toxicity (sections 1.1.2, 1.2.1). The draft assessment concludes that male and female reproductive effects are a human hazard of benzo[a]pyrene exposure. Do the available human, animal and mechanistic studies support this conclusion?

2c. Immunotoxicity (sections 1.1.3, 1.2.1). The draft assessment concludes that immunotoxicity is a potential human hazard of benzo[a]pyrene exposure. Do the available human, animal and mechanistic studies support this conclusion?

2d. Cancer (sections 1.1.5, 1.2.2). The draft assessment concludes that benzo[a]pyrene is "carcinogenic to humans" by all routes of exposure. Do the available human, animal, and mechanistic studies support this conclusion?

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?

Dose-Response Assessment (Section 2)

Question #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 (RfD): The draft assessment proposes an overall reference dose of 3×10^{-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?

Preliminary Comments:

- *The discussion around identification of studies and effects used in the dose-response analysis was clear and to the point. Analysis is consistent with EPA's Benchmark Dose Technical Guidance Document (U.S. EPA 2012c). Adequate information is provided in Appendix E of the Supplemental Information to assess model fit adequacy.*
- *A review of model fits found no situations where the chosen model looked inadequate.*
- *It seems appropriate to establish a P-value upper threshold of 0.01 to indicate inadequately fitting models allowing discarding of marginally adequate models and allowing the focus to remain on the best fitting models.*
- *The model selection protocol tends to favor PODs that are conservatively low. When multiple models produce BMDL estimates within 3x of each other, the model AIC criteria is used to select the value to be used, whereas when the BMDL estimates are not within 3x, the lowest BMDL is used. This (standard) selection protocol results in a value that could have very conservative properties.*

3b. Inhalation Reference Concentration (RfC): The draft assessment proposes an overall reference concentration of 2×10^{-6} mg/m³ based on decreased fetal survival during a critical window of development (Section 2.2). 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 2.2.5) reflect the scientific considerations that are inherent for exposures during a critical window of development?

3c: Oral slope factor for cancer (section 2.3). The draft assessment proposes an oral slope factor of 1 per mg/kg-d based on alimentary tract tumors in mice. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?

Preliminary Comments:

- *This section presents the oral slope factor case in a straightforward discussion, providing adequate justification for decisions at each step.*
- *The assumption of equal cumulative exposure yielding equivalent outcomes (page 2-18, lines 18-21) - used to justify converting administered dose (5days/week) to equivalent continuous dose (7days.week) – is not discussed in either the report or in the Supplement. Is this a generally accepted assumption? Is this the only method of converting administered dose to equivalent continuous dose? Originally I wondered whether the PBPK animal models (Appendix D, section D.2) might offer additional approaches, but the models have high uncertainties and model a limited number of the important pathways and or sources/sinks that limit their utility.*

- *Uncertainties in the Derivation of the Oral Slope Factor - Use of Multistage Weibull model is well supported – Desire to incorporate time of death in the modeling is appropriate and would be expected to improve model fits. The analysis presented seems to represent a due-diligence effort that is supported by reviews of this model and approach by previous expert panels. This does not mean that multistage Weibull is the perfect model or that some other model incorporating time of death could be found to fit better. But, there do not seem to be other obvious models that could have been fit that would incorporate the available data.*
- *Statistical uncertainty is appropriately accounted for and use of BMDL justified.*
- *Previous IRIS Assessment Oral Slope Factor - The major difference with the previous IRIS assessment seems to be due in a large part to the use of $BW^{2/3}$ scaling instead of the $BW^{3/4}$ scaling used in this assessment.*

3d: Inhalation Unit Risk: The draft assessment proposes an inhalation unit risk of 0.6 per mg/m³ based on a combination of several types of benign and malignant tumors in hamsters. Is this value scientifically supported, giving due consideration to the intermediate steps of selecting studies appropriate for dose-response analysis and calculating points of departure?

Preliminary Comments:

- *This section presents the approach to estimation of the inhalation unit risk case in a straightforward way, providing adequate justification for decisions at each step. The approach used (accommodating time-to-tumor data in a Multistage Weibull model) is similar to that used to estimate the oral slope factor. Model fits look adequate for obtaining reasonable estimates.*
- *The analysis proceeds by considering tumors to be either all fatals or all incidentals. This seems reasonable given the absence of investigator-determined cause of death and the likelihood that all tumors are unlikely to be fatal.*
- *Not discussed in Table 2-10 (uncertainties).*
 - *The impact on estimates of assuming something other than equal risk for all species is associated with equal concentrations in air (page 2-35, lines 4-5).*
 - *The impact on estimates of assuming something other than equal cumulative exposure yields equivalent outcomes (page 2-18, lines 18-21- used to justify converting administered dose (5days/week) to equivalent continuous dose (7days.week)) This is not discussed in either the report or in the Supplement. Is this a generally accepted assumption? Is this the only method of converting administered dose to equivalent continuous dose? Originally I wondered whether the PBPK animal models (Appendix D, section D.2) might offer additional approaches, but the models have high uncertainties and model a limited number of the important pathways and or sources/sinks that limit their utility.*

- *The impact on estimates of assuming that the latency time, t_0 (the time between a tumor first becoming observable and causing death) is different from zero (page E-66, line 23-24).*
- *The impact on estimates of eliminating from the analysis all animals without confirmation of one or more of the pharynx or respiratory tract tissues being examined, unless a tumor was diagnosed in those that were examined (page E-66, lines 25-27). This decision impacts the denominator of the cases fraction and hence has the potential to impact significantly the model results. On the other hand, sample sizes are quite good and if not a lot of cases are excluded the impact could be small. **Note:** In Appendix G (page G-8, lines 43-46) we are informed that 5 low-exposure animals are omitted in the dose response modeling. Including this information directly into the Report or even into the body of the Supplement and discussing it briefly would resolve this comment.*

3e: Dermal Slope Factor for cancer: The draft assessment proposes a dermal slope factor of 0.006 per ug/day based on skin tumors in mice. 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 scaling from mice to humans? Does the method for cross-species scaling (section 2.5.4 and appendix E) reflect the appropriate scientific considerations?

Preliminary Comments:

- *The approach to estimation of the dermal slope factor is presented in a straightforward manner*
- *There is more dermal studies, with more dose levels within studies and more than adequate animals/dose. This resulted in EPA having to fit multiple models for each dataset and then select among multiple model forms for those best fitting. As a result the modeling section in Appendix E is larger and harder to follow. The modeling results presented in Appendix E do allow following and duplicating the EPA analysis feasible.*
- *Cannot comment on the method used for interspecies scaling of the dermal slope factor, but this is an important question because it contributes quite significantly to the uncertainty in the final estimate (see also comment and answers on page G-12, lines 5-7).*

3f: Age-dependent adjustment factors for cancer (section 2.6). The draft assessment proposes the application of age-dependent adjustment factors based on a determination that benzo[a]pyrene induces cancer through a mutagenic mode of action (see the mode-of-action analysis in section 1.1.5). Do the available mechanistic studies in humans and animals support a mutagenic mode of action for cancer induced by benzo[a]pyrene?

Questions #4: Executive summary. Does the executive summary clearly and appropriately present the major conclusions of the assessment?

Preliminary Comments:

- *The Executive Summary is 6 and ½ pages in length – long for a summary. A lot of the text is duplicated from the body of the Report. Including only the table information (combined into one long table including cancer slope factors and estimates using ADAFs – add uncertainty factors where utilized) followed by a list of notes – one per table row in bullet format, I estimate the summary could be just as understandable in about 3 pages. Additional bullet notes could be added to summarize estimation method issues (eg. Use of Multistage Weibull and Multistage Cancer models to accommodate time-to-tumor data in estimating cancer slope factors.) A reduced Executive Summary would further meet the spirit of NRC recommendation 1.*
- *An Executive Summary should be able to stand-alone and not make references back to the body of the Report or to the Supplement (see page xxxvi, line 19).*

Question #5: Charge Question on 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

Preliminary Comments:

- *The EPA response to the comment on “Apparent threshold in animal cancer bioassays” (page G-7, line 7-25) is in line with good statistical practice/thinking. The argument for or against an exposure threshold below which cancer effects might not occur must be answered through increased biological processes understanding because there are simply not enough test animals available to answer the question experimentally and through empirical models.*
- *The EPA response to the comment on “Exposure variability in the study used to derive inhalation unit risk” (page G-9, lines 1-14) provides an answer that essentially “begs the question”. Agreed, the way EPA “eliminates” this issue is by assuming that cancer risk is proportional to cumulative exposure, but still the impact of exposure variability should be better addressed in the body of the Supplement.*

Other Comments that do not necessarily fit in any one section:

- *Page 2-13, line 16-17 states: "...and the study used to derive a candidate value based on decreased testosterone (Zheng et al., 2010) did not observe a dose-response relationship (a 15% decrease in testosterone was seen at the low and high doses, with statistical significance at the high dose)." The study did not observe a monotonic dose-response pattern as might be expected for a toxin, but the inverse U shaped response pattern is not discussed further. My limited understanding of such patterns is that they may be suggestive of endocrine disruption. This issue is not discussed. I did note that the literature review key words did include endocrine disruption but none of the papers referenced in the report or the supplement discuss endocrine disruption.*
- *Good point made on page 2-14, lines 25-28 regarding exposure level at critical development windows. Need more be said? The discussion on ADAF in, section 2.6 does not come back to this critical developmental life stage discussion.*
- *In a few places in the Report and/or the Summary, reference was made to historical incidence rates of outcomes. All the multistage models assume a background factor in estimating the response likelihood. The key data allowing estimation of this parameter is the control animals from each study. For many of the test species, background incidence rates for tumors can also be estimated from historical control datasets. Incorporating historical control information into the modeling process conceptually can be done and would reduce the uncertainty of the response at the 0 dose and could as a result reduce the overall uncertainty in the model and the BMDL specifically. I have not seen this done anywhere but wondered if this was attempted for any of the exposure scenarios presented. The Report refers to the potential for using historical controls (see page xx, lines 30-41) but I can't find where this was actually done in this Report.*

Dr. Steve Roberts

2.1.2 Methods of Analysis, Dosimetric Adjustment Factor

PODs estimated based on effects in adult animals were converted to human equivalent doses (HEDs) employing a Dosimetric Adjustment Factor consistent with current EPA guidance. $BW^{3/4}$ scaling was not applied to calculate HEDs from studies in which doses were administered to early postnatal animals, which I think is also consistent with EPA guidance [will confirm]. I have no comments or suggestions for improvement.

2.3.1 Analysis of carcinogenicity data (Choice of studies)

Available studies are identified and the rationale for selection of Kroese et al. (2001) and Beland and Culp (1998) as the best studies is presented clearly and concisely. I agree with the selection of these studies as the most appropriate for conducting dose-response analysis.

2.5.1 Analysis of carcinogenicity data (Choice of studies)

Available studies are identified and described succinctly, with details presented in the Appendix. The rationale for selection of the NIOSH study as the best for dose-response analysis is presented clearly and concisely. I agree with the selection of this study.

2.5.2 Dose-response analysis

It is clear from the discussion in this section and in Appendix E that the appropriate dose metric has not been established for skin tumorigenicity from benzo(a)pyrene. Appendix E discusses some options for dose-metrics, but this discussion is in the context of how to extrapolate observations in mice to humans and seems to ignore the basic question of what is the appropriate metric in *any* species (mass; mass per unit area; something else?). This is not just an issue for extrapolation among species, but also for the fundamental form that dermal cancer potency factor should take. This question could be addressed experimentally, but to my knowledge has not to date. Until this issue is resolved, developing a dermal slope factor is premature, in my opinion.

Executive Summary

In general, the Executive Summary clearly presents the major decisions and conclusions of the assessment. The extent to which the committee agrees with those decisions and conclusions will be determined during the face-to-face meeting. There will be some suggestions for improvement; for example, under "Key Issues Addressed in Assessment" the issue of applicability of the overall RfD and RfC values to risk assessment for the general population is raised, which is an important one. Unfortunately, the Executive Summary says nothing about it other than to refer the reader to two sections in the main body of the report. The Executive Summary should bring forward at least the main ideas regarding this topic.

Parenthetically, the Executive Summary makes statements such as, "These organ- or system-specific reference values may be useful for subsequent cumulative risk assessments that consider the combined effect of multiple agents acting at a common site." The statement is accurate, but the reference values as presented are arguably unsuitable for cumulative risk assessment without some modification. [The issue is that the UF_D one would apply to produce a candidate reference value that might be selected as the overall reference value, as in this report, could be different from the UF_D one would apply to create a value that applies to one specific type of toxicity, e.g.,

hepatotoxicity, in a cumulative risk assessment). An explanation of this issue may not be appropriate for the toxicological review, but should be articulated by EPA *somewhere*. The current draft TMB review suggests that EPA provide some discussion and/or guidance on this subject. This committee may want to decide whether to bother reiterating that suggestion.

Summary of External Peer Review

It will be important to make clear in our response to this charge question that we are considering only public comment summaries presented in Appendix G. Without seeing the original public comments, we have no way of knowing how accurately the comments were summarized and whether the points made by the public commenters were adequately captured.

EPA accepted some public comments and made suggested changes while rejecting others. When the EPA disagreed with a comment, a clear explanation of the basis for disagreement was provided. Our charge includes determining “whether there are scientific issues that were raised by the public as described in Appendix G that may not have been adequately addressed.”

Presumably this means situations where the committee either disagrees with the EPA response or thinks that they missed the point. Some (hopefully most) of these situations will become clear as the earlier charge questions are discussed by the committee. Points on which there is disagreement between the EPA and public commenters that are not addressed previously by the committee will need to be discussed under this charge question. Points potentially include: [my initial impressions in brackets]

- Low confidence placed by the EPA in skin cancer studies using mice with human skin grafts [EPA makes some good points on limitations of these studies]
- Disagreement by EPA with public comments indicating that coal tar studies in humans demonstrate that benzo(a)pyrene does not cause skin cancer in humans [EPA makes some good points on limitations of these studies]
- Contention that EPA has mischaracterized the evidence supporting an association between benzo(a)pyrene exposure and lung and skin cancer in humans [lung cancer association looks solid to me, but the association with skin cancer might be overstated]
- Why is “decreased anxiety” a critical effect? [who wouldn't want decreased anxiety? ... but it does comport with EPA guidance]
- Disagreement by EPA with public comments that dose-response modeling of data from cancer bioassays from oral, inhalation, and dermal routes show thresholds. [Agree with EPA that this type of modeling cannot identify thresholds for carcinogenesis.]
- Disagreement by EPA with public comments indicating that the Thyssen et al. (1981) inhalation study was unsuitable for development of an inhalation unit risk because the maximum tolerated dose was exceeded and exposures were highly variable over time. [Still thinking about this one.]
- Disagreement by EPA with public comments indicating that studies show non-linear dose-response relationships for skin cancer and benzo(a)pyrene and there is a MOA based upon inflammation, cell killing, and cell replication, consistent with non-linearity. [I concur with EPA's response on this one]
- Disagreement by EPA with public comments indicating that PAHs are not casually related to human skin cancers because PAH-induced tumors in mouse skin have a different genetic signature than human skin tumors. [Generally concur with the EPA response]

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

- Disagreement by EPA with comments that the study by Sivak et al. (1997) should not be used to develop a dermal slope factor because the maximally tolerated dose was exceeded. [EPA response seems adequate.]
- Disagreement regarding EPA's approach for extrapolating the dermal slope factor from mouse to human skin and with the expression of the slope factor in $\mu\text{g}/\text{d}$ [I also disagree with EPA's approach. This will no doubt be discussed in the context of other charge questions.]
- Disagreement over the risks that would be estimated for the general population based upon the proposed dermal slope factor and typical PAH exposures. [I have not yet gone through the math.]

Dr. Richard Schlesinger

Charge Question 3b:

Based upon the available database, the appropriate critical toxicological effects for estimation of an RfC are developmental and reproductive parameters. This choice of endpoints is supported by consistent qualitative effects from exposure across species. Unfortunately, there are very limited datasets for evaluation of chronic effects in this regard from inhalation exposure. The specific key study used in the Assessment for derivation of the RfC, namely that of Archibong et al (2002), employed three concentrations of B(a)P and examined decreased fetal survival; since the lowest level used still resulted in toxic effects, this was considered as the LOAEL for the POD for dose-response analysis.

p. 2-18. L. 1-18. The rationale for use of a value of 860 mL for tidal volume (TV) and 50 mL for volume of the upper respiratory tract (URT) is not clear. On the average, TV ranges from 7-9 mL/kg BW, so for a 70 kg person (the default body wt for human), the value should range from 490-630 mL. Using the value of 16bpm noted in the B(a)P Document, the minute volume would range from 7.8-10.1 L/min. Regarding FRC, which is RV + ERV, the average value in males is 2400 mL, so the value in the Document of 3,300 mL seems high.

p. 2-19. Section 2.2.3. This section discusses the rationale for the various UFs used. What is not clear is the rationale for the UF used for subchronic to chronic extrapolation, especially since the key study for derivation of RfC was subacute. While clearly pregnancy is not a chronic condition, animals could have been exposed over a lifetime with multiple pregnancies and perhaps the exposure level for a toxic effect may have been lower than those in the study. The B(a)P Document indicates a UF of 1; the default in this case is 10, and this may have been a better value to use since the inherent extrapolation in the Assessment involves subacute to chronic. Furthermore, the UF for interspecies differences should be raised to 10 for several reasons, including the fact that the particle size of the aerosol used in the key study had about 45% of the mass in particles above 2 μ m, which would deposit in the upper respiratory tract of rats but would deposit in the deeper lung in humans. Thus, since responses and toxic effects are dependent upon regional deposition, the UF should be raised.

However, the Methods for Derivation of RfC as noted in the 2002 Document cited above is that the UF total should be less than or equal to 3000, and a full value of 10 should not be used in 4 or more areas of extrapolation. The current B(a)P Assessment fits this criterion using 1 for the subchronic to chronic, but would not if using the more appropriate value of 10.

p.2-20, L. 9. What is meant by, "...these studies observed a high magnitude of response"?

p.2-20, L. 30-32. This statement seems to contradict the selection of the POD for the RfC.

p.20-23, Section 2.2.6. Confidence in the derived value for RfC is noted as low to medium. Confidence in the database is indicated as low for a number of reasons, while confidence in the key study is noted as medium. The rationale for the range of confidence up to medium for the derived RfC value is noted as due to "...consistent systemic effects observed by the oral route...and similar effects observed in human populations exposed to PAH mixtures." However, while there is consistency in qualitative effects between

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

oral and inhalation routes of exposure, there may not be consistency in the dose-response relationship between different routes of exposure. Furthermore, effects of B(a)P do occur at site of entry, so some effects will differ between oral and inhalation exposure. Thus, using the rationale above to increase confidence of the RfC value to medium is questionable.

On page 2-25, lines 1-5, it is noted that the study selected as the basis for the RfC "...provided limited information regarding the inhalation exposures of the animals..." in that it was not clear whether the concentrations noted in the paper were target values or analytical concentrations nor was the "...method used to quantify benzo(a)pyrene in the generated aerosols..." reported. This, together with the comments above and the magnitude of the UFs used, indicate that the overall RfC confidence should be low, rather than low to medium. In fact, the Methods for Derivation of RfC document notes, "Low confidence in an RfC is usually applied to a derivation that is based on several extrapolations and indicates an estimate that may be especially vulnerable to change if additional data become available. For some chemicals, the data base is so weak that the derivation of a low confidence RfC is not possible." I think the current case is the former rather than the latter.

The preference for a POD for the RfC is a NOAEL, but the study used for the B(a)P value is a "default" LOAEL based upon the lowest of a number of discrete concentrations used in the key study rather than an extensive dose-response relationship. Thus, the actual "true" LOAEL is not clear, and necessitated use of UFs to compensate.

Finally, in the Methods for Derivation of RfC, the availability of only one inhalation bioassay is noted as the minimum database for estimation of an RfC and the confidence will then be low. Furthermore, the derivation procedure notes that in this case a chronic study is preferred, but a subchronic is acceptable. In the B(a)P RfC derivation, it is a subacute study (Archibong et al. 2002) that is being used for POD. Thus, all of the above strongly suggests a confidence level of low for the derived RfC, and not low to medium.

Charge Question 3d:

Page 2-35, L. 26-28. What is the basis for this statement and how useful is the derived unit risk in this case. How does the value proposed compare with actual exposure estimates of humans.

Page 2-36, Section 2.4.4. A number of uncertainties are discussed, but what is the overall uncertainty or level of confidence for the number derived.

Dr. Leslie Stayner

1. **Literature search/study selection.** Is the literature search strategy well documented? Please identify additional peer-reviewed studies that might have been missed.

The literature search strategy was thorough and very well documented in the toxicological review. Nonetheless, it appears that the epidemiologic literature presented in the toxicological review and supplemental information regarding the carcinogenicity of BAP was incomplete and somewhat out of date.

The toxicological review emphasized studies that met their criteria for high quality (i.e. Tier 1). Although I agree with most of the criteria they chose for identifying high quality studies, I believe that requiring a detailed exposure assessment for BAP is unnecessarily restrictive. Studies with a detailed exposure assessment would be most relevant for an exposure-response assessment, but are not necessary for hazard identification. The review only considered that three studies met their criteria for Tier 1 for lung cancer (Armstrong and Gibbs 2009, Spinelli et al 2006, Xu et al. 1996) and bladder cancer (Gibbs and Sevigny 2007a and 2007b, Spinelli et al 2006, Burstyn et al 2007). The Tier 1 studies only included studies of the aluminum and iron and steel manufacturing. It did not include any studies of workers from the coke ovens, roofing or asphalt industries which would have very high exposures to BAP and thus should be relevant for determining causality even though they may not have had detailed exposure assessments for BAP. Tier 2 studies are presented in a table in the report. However, there are many studies missing from these tables (e.g. Romunstadt et al. 2000, Ronneberg 1999, that have been included in prior reviews (i.e. see Table 1 in Bosetti et al. 2007, and Rota et al. 2014).

There is a disconnect between the review presented in the toxicological review and the supplemental information section. Normally I would expect a supplement to provide additional information then what is presented in the main body of the report. However, the supplemental information section did not follow the same logic of reviewing Tier 1 and Tier 2 studies, and did not provide a more detailed review of the studies then the main report.

The review presented in the supplemental information section relied heavily on a systematic review and meta-analysis reported by Bosetti et al. in 2007, and by Armstrong et al. in 2004. It seems inappropriate for EPA to rely solely on a review articles rather than a review of the primary literature. There is also a more recent meta-analysis that they did not include in their review (Rota et al. 2014). Many of the epidemiologic studies cited in Bosetti and Rota are not discussed in the EPA supplemental document. For aluminum production workers the EPA only discusses the studies by Spinelli et al. (1991, 2006), Romundstad et al. (2000a and 2000b) and Xu et al. (1996). There are 10 other studies of aluminum production workers cited in the Bosetti review (see Table 1 of Bosetti), and five additional studies cited in the Rota review article (see Table 1 of Rota). It is unclear why the EPA only included the few studies that they did review in their report.

The EPA supplemental review doesn't discuss any of the studies of asphalt workers and roofers or coke oven workers. For asphalt and roofers they refer the readers to the Bosetti et al (2007)

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

review, which as mentioned above, was updated by Rota et al. (2014). They cite five papers as providing evidence of an excess risk of lung cancer and weak evidence for bladder cancer among asphalt workers and roofers (Burstyn 2007, Partanen 1994, Chiazze 1991, Hansen 1991, 1989, and Hammond 1976). They seemed to have overlooked studies cited in Bosetti (see Table 1) of roofers by Swaen et al., and of asphalt workers cited in Rota (see Table 1) by Behrens et al. (2009) and Zanardi et al. (2013). In addition given the differences in the nature of exposure, I think that it would be appropriate to separately discuss the findings for asphalt workers and roofers.

For coke oven workers, coal gasification and iron and steel foundry workers the supplemental report relies entirely on the reviews by Boffetta et al. (1997), Bosetti et al. (2007) and Armstrong et al. (2004). The more recent review by Rota (2014) identified two new studies of iron and steel workers (see Table 1) that were not considered in the earlier reviews. I was able to identify one additional study of coke oven workers by Miller et al. (2013), which was not included in the Rota or EPA reviews.

Finally, it is not clear why some of the studies of coal tar that were identified in the comments from the American Coke and Coal industry were not included in the EPA review. In particular the studies by Bhate et al (1993), Hannuksela-Svahn et al (2000), Jemec, G.B.E. and A. Østerlind (1994), Jones S.K. et al (1985), Menter A. and D.L. Cram (1983), and Muller and Kierland (1964) seem relevant.

2e. **Cancer** (sections 1.1.5, 1.2.2). The draft assessment concludes that benzo[a]pyrene is “carcinogenic to humans” by all routes of exposure. Do the available human, animal, and mechanistic studies support this conclusion?

The epidemiologic data alone is not sufficient to conclude that BAP is carcinogenic to humans. There is strong evidence that workers in industries with high exposures to BAP are at increased risk of lung and to a lesser extent for bladder cancer. However, workers in these industries are all exposed to other PAHs and it was impossible to single out BAP in the analyses of these studies. However, the epidemiologic data combined with the animal and mechanistic data do provide strong support for the conclusion that was reached by the EPA that BAP is carcinogenic to humans.

References

Armstrong BG, Gibbs G (2009) Exposure-response relationship between lung cancer and polycyclic aromatic hydrocarbons (PAHs). *Occup Environ Med* 66:740–746.

Armstrong B, Hutchinson E, Unwin J, Fletcher T. Lung cancer risk after exposure to polycyclic aromatic hydrocarbons: a review and meta-analysis. *Environ Health Perspect.* 2004 Jun;112(9):970-8.

Behrens T, Schill W, Ahrens W (2009) Elevated cancer mortality in a German cohort of bitumen workers: extended follow-up through 2004. *J Occup Environ Hyg* 6:555–561.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Bosetti C, Boffetta P, La Vecchia C. Occupational exposures to polycyclic aromatic hydrocarbons, and respiratory and urinary tract cancers: a quantitative review to 2005. *Ann Oncol*. 2007 Mar;18(3):431-46.

Boffetta P, Jourenkova N, Gustavsson P. Cancer risk from occupational and environmental exposure to polycyclic aromatic hydrocarbons. *Cancer Causes Control* 1997;8:444-472.

Bostrom CE, Gerde P, Hanberg A, et al. Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. *Environ Health Perspect* 2002;110:451-488.

Gibbs GW, Sevigny M (2007a) Mortality and cancer experience of Quebec aluminum reduction plant workers, part 4: cancer incidence. *J Occup Environ Med* 49:1351-1366.

Gibbs GW, Sevigny M (2007b) Mortality and cancer experience of Quebec aluminum reduction plant workers. Part 3: monitoring the mortality of workers first employed after January 1, 1950. *J Occup Environ Med* 49:1269-1287.

Miller BG, Doust E, Cherrie JW, Hurley JF. Lung cancer mortality and exposure to polycyclic aromatic hydrocarbons in British coke oven workers. *BMC Public Health*. 2013 Oct 16;13:962.

Romundstad P, Andersen A, Haldorsen T. Cancer incidence among workers in six Norwegian aluminum plants. *Scand J Work Environ Health* 2000;26:461-469.

Ronneberg A, Haldorsen T, Romundstad P, Andersen A. Occupational exposure and cancer incidence among workers from an aluminum smelter in western Norway. *Scand J Work Environ Health* 1999;25:207-214.

Rota M, Bosetti C, Boccia S, Boffetta P, La Vecchia C. Occupational exposures to polycyclic aromatic hydrocarbons and respiratory and urinary tract cancers: an updated systematic review and a meta-analysis to 2014. *Arch Toxicol*. 2014 Aug;88(8):1479-90.

Spinelli JJ, Band PR, Svirchev LM, Gallagher RP. Mortality and cancer incidence in aluminum reduction plant workers. *J Occup Med* 1991;33:1150-1155.

Swaen GMH, Slangen JJM, Volovics A, et al. Mortality of coke plant workers in the Netherlands. *Br J Ind Med* 1991;48:130-135.

Xu Z, Brown LM, Pan GW, et al. Cancer risks among iron and steel workers in Anshan, China, Part II: Case-control studies of lung and stomach cancer. *Am J Ind Med* 1996;30:7-15.
Zanardi F, Salvarani R, Cooke RM, Pirastu R, Baccini M, Christiani D, Curti S, Risi A, Barbieri A, Barbieri G, Mattioli S, Violante FS (2013) Carcinoma of the pharynx and tonsils in an occupational cohort of asphalt workers. *Epidemiology* 24:100-103.

Dr. Alan Stern

Comments on Charge Question 2d

Cancer (sections 1.1.5, 1.2.2). The draft assessment concludes that benzo[a]pyrene is "carcinogenic to humans" by all routes of exposure. Do the available human, animal, and mechanistic studies support this conclusion?

The evidence for BaP carcinogenicity by all routes of exposure in animal models is strong, and EPA has done a good job of presenting these animal data. Given the global potential for BaP to cause cancer in animals, it is also highly likely that BaP can also cause cancer in humans by all routes of exposure. However, at least two requirements for assignment of the category of "carcinogenic to humans" in EPA's 2005 Guidelines for Risk Assessment present logical problems relative to making this assignment for BaP based on the arguments presented in the draft IRIS document.

The first, is "(a) there is strong evidence of an association between human exposure and either cancer or the key precursor events of the agent's mode of action but not enough for a causal association." With respect to this requirement, there does not appear to be any epidemiologic evidence that addresses exposure to BaP in isolation from other PAHs, at least some of which are also known animal carcinogens. EPA presents an argument on pg. 1-83, lines 26-29 that "...the exposure-response patterns seen with the BaP measures make it unlikely that these results represent confounding by other exposures." However, the specific evidence behind this statement is not clear. With respect to dermal carcinogenicity in humans, there is likewise no evidence from BaP-specific exposures. In this case, EPA's argument (pg. 1-84, lines 3-7) is based on the relationship between benzo[a]pyrene diol epoxide-DNA (BDE-DNA) adducts as both a marker of BaP exposure and a causal step in skin tumor production. However, while it seems clear that BDE-DNA adducts are a marker of BaP exposure, the role of BDE-DNA adducts as a necessary step in BaP carcinogenicity is less clear. For example, on pg. 1-74, lines 30-38, EPA in its review of Culp et al. (1996) notes that in mice exposed to BaP in the diet had a sharp increase in tumor response between the lowest and next highest dose while the BDE-adduct concentration increased linearly. EPA's argument that BaP meets this first criterion for assignment of the "carcinogenic to humans" category may hinge on the "...but not enough for a causal association" portion of this criterion (although EPA does not explicitly say so, this requirement appears to function when the first, stand-alone requirement for the assignment of "carcinogenic to humans": "... *convincing epidemiologic evidence of a causal association between human exposure and cancer*" is not met). Clearly, this requires interpretation of specificity of this requirement, but I think that EPA could do a better job of making this case.

The second logical problem relates to the linked requirement in the 2005 guidelines that "(d) there is strong evidence that the key precursor events that precede the cancer response in animals are anticipated to occur in humans and progress to tumors, based on available biological information." In the IRIS document's discussion of carcinogenic modes of action, EPA presents three possible (and non-mutually exclusive) modes of action (pg. 1-69), BDE-DNA adducts,

radical cation, and *o*-quinone-ROS. As discussed above, notwithstanding the qualitative carcinogenic potential of BDE, the quantitative relationship between BDE-DNA and tumor production does not appear to be clear and if possible, EPA should present a synthesis of evidence in animals and/or humans that supports a dose-response relationship for BDE-DNA and tumors. As for the other two possible modes of action, it appears that the evidence supporting these modes of action is from animal and/or *in vitro* studies. It may be sufficient for EPA to make the case that these are both key-precursor events and highly likely to function in humans given the basic underlying biochemistry and molecular biology in mammals. However, the current document does not explicitly make this case.

The above notwithstanding, I want to make it clear that these considerations relate only to evidence requirements for the assignment of the specific category of "carcinogenic to humans" and not to the highly likely carcinogenic potential for BaP to humans.

Comments on Charge Question 3a (section 2.1.3 – Uncertainty Factors – RfD)

EPA's discussion of uncertainty factors (UFs) begins with the UF_H . Given that the starting point for this process is the animal-based POD, the UF_H is not a logical place to begin this discussion. Rather, the discussion should begin with the uncertainty factor for LOAEL-NOAEL conversion (UF_L), followed by the UF for subchronic-chronic conversion (UF_S), then the UF for animal-human (UF_A) and *then* proceed to the UF_H .

On pg. 2-9, lines 5-10, EPA provides the rationale for applying a UF of 3 (as opposed to the full standard UF of 10) for accounting for potential differences between animals and humans. The reduction of the UF to 3 by the prior application of $bw^{3/4}$ allometric scaling is consistent with EPA's risk assessment guidance. However, the rationale presented in the document at this point, that the $bw^{3/4}$ accounts for aspects of toxicodynamic as well as toxicokinetic differences between animals and humans, does not appear to be consistent with EPA guidance, nor does it make toxicological sense to me. What is the basis for stating that allometric scaling necessarily addresses interspecies toxicodynamic differences? Toxicodynamic differences would conceptually appear to arise from differences in genetics and biochemistry and these factors should not necessarily scale as a function of body weight. If EPA believes otherwise, the document should either cite prior EPA guidance to this effect, or provide a more detailed basis for this statement. The issue here is not the value *per se* of this UF (with which I agree), but with its explanation.

The application and justification of the other UF for non-cancer oral dose endpoints appears reasonable and consistent with EPA guidance.

Comments on Charge Question 3b (section 2.2.3 – Uncertainty Factors – RfC)

As per my comments on the presentation of UFs for the RfD, EPA's discussion of uncertainty factors for the RfC begins with the UF_H . Given that the starting point for this process is the animal-based POD, the UF_H is not a logical place to begin this discussion. Rather, the discussion should begin with the uncertainty factor for LOAEL-NOAEL conversion (UF_L), followed by the

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

UF for subchronic-chronic conversion (UF_S), then the UF for animal-human (UF_A) and *then* proceed to the UF_H.

A UF_A of 3 was chosen to account for uncertainty in extrapolation from animals (rats) to humans. The rationale for not applying the full UF of 10 was the application of the regional deposition dose ratio (RDDR) based on EPA's 1994 guidance. The text states that the application of a dose adjustment factor (DAF) derived using this methodology accounts for interspecies toxicokinetic differences and therefore, only the residual uncertainty in the potential interspecies toxicodynamic differences remain to be addressed by the application of a UF. However, the RDDR only addresses interspecies differences in particle deposition. Since the developmental and reproductive PODs to which this UF are applied are systemic in nature (i.e., result from effects occurring outside the respiratory tract subsequent to absorption from the respiratory tract), interspecies differences in particle deposition do not account for interspecies differences in toxicokinetics subsequent to respiratory deposition. Unlike the estimation of the human equivalent dose (HED) in the derivation of the RfD, where toxicokinetic differences are addressed by a $bw^{3/4}$ adjustment, the use of only the RDDR implicitly assumes a body weight adjustment on a linear basis and the text does not appear to address any other body weight-based adjustments. This would imply that some interspecies toxicodynamic uncertainty remains. It is not clear that this necessarily requires the application of the full UF_A of 10, however, this residual toxicokinetic uncertainty should be addressed.

The application of the remaining UFs appears to be consistent with EPA guidance.

Comments on Charge Question 3f (section 2.6 – Age Dependent Adjustment Factor)

The discussion in section 1.1.5 presents three strong lines of evidence for a mutagenic mode of action (or several mutagenic modes of action) for BaP: The observation that BaP is a complete carcinogen in skin painting studies; the production of DNA base transversions resulting from benzo[a]pyrene diol epoxide (BDE) adducts, and the consistently positive results in bacterial mutagenicity models (with metabolic activation). Two other carcinogenic mechanisms, radical cation production and *o*-quinone/ROS production, although consistent with mutagenicity, provide weaker evidence for a mutagenic mode of action, as they are also consistent with non-mutagenic modes of action. I believe that this evidence makes a plausible case that BaP can cause cancer through a mutagenic mode of action and hence, justifies the application of the Age Dependent Adjustment Factor (ADAF). However, this case would be strengthened if EPA could provide evidence as to the relative contribution of each of these mechanisms to overall cancer risk showing that those mechanisms resulting in mutagenicity predominate or at least can be assumed to account for a large portion of the modeled tumorigenicity. Such evidence could take the form of (e.g.) the relative production and potency of BDE adducts compared to radical cation production and *o*-quinone/ROS production at relevant doses of BaP, and/or the relative kinetics and half-lives in the nucleus of these proximate carcinogenic agents.

Dr. Charles Vorhees

Charge Question #1

Literature search/study selection. Is the literature search strategy well documented? Please identify additional peer-reviewed studies that might have been missed.

RESPONSE:

The EPA literature search is thorough, well-documented, and comprehensive. One newer reference for the human data is provided below that the EPA may want to consider.

Perera et al. (2014) on PAH exposure and ADHD in children (Perera et al., 2014).

There are two experiments in animals cited by the EPA but not entirely in each context where the data may be of value. One is by Patri et al. on BaP in developing rats on learning and the role of norepinephrine as a potential protective factor against BaP-induced neurotoxicity (Patri et al., 2013). The other is on BaP in rats on motor and cognitive behavior (Maciel et al., 2014) that is partially relevant to the developmental neurobehavioral effects inasmuch as it supports the data that BaP is neurotoxic at different stages of ages of development.

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 and animal studies support this conclusion?

RESPONSE:

The Committee subdivided this Charge Question into two parts: 2a-1 on developmental neurotoxicity and 2a-2 on developmental toxicity other than neurodevelopment.

For Charge Question 2a-1: Developmental Neurotoxicity the EPA report was thorough with no significant missed literature in the view of the committee (see above).

Human Studies: There are a series of relevant human epidemiological studies on developmental BaP-PAH effects on neurodevelopment, including, but not limited to: (Perera et al., 2012b; Perera

et al., 2011; Tang et al., 2006; Perera et al., 2005; Perera et al., 2004; Tang et al., 2008; Perera et al., 2012a; Perera et al., 2009). These support the EPA's view that BaP is a human developmental

neurotoxic agent. The EPA appropriately notes that in human studies the exposures are to PAH mixtures and, therefore, the individual effects of BaP alone on child neurodevelopment cannot be isolated and determined to be exclusively attributable to BaP rather than the sum, interaction, or even antagonist effect of multiple PAHs acting in concert. However, the human cohort studies have many strengths. These include that they are in the target species: human, that they are prospective studies, and they are from two separate populations: (1) in New York City with investigators from Columbia University in which the cohort was identified and followed from before birth to the age of 9 years so far with ongoing follow-up evaluations extending to the future. This study features consistent reassessments at regular intervals, the use of well-standardized human neuropsychological instruments administered by a experienced psychometrists, high level subject retention over the many years, and careful statistical analyses of the data controlling for possible cofactors; and (2) a separate prospective cohort study conducted in collaboration with the Columbia-NYC investigators in Tongliang, China by Tang et al. This study is also a prospectively enrolled developmental cohort study and the data stemming from it are in agreement with the findings from the Columbia-NYC study, lending further credence to the effects of PAH exposure on children's neuropsychological and physical development. An important aspect of the human studies that add additional weight to their validity is that they measured BaP-specific DNA adducts in maternal and umbilical plasma and also used individually worn air samplers on the mothers and found general agreement between the air sampling and internal dose metrics. Of importance is that the method used for the BaP DNA adduct determinations was specific for BaP adducts and not generic for other PAH DNA adducts. That the Columbia-NYC study used a specific DNA adduct assay for BaP is a significant strength of these data.

Animal Studies: The animal experiments on effects of BaP on the nervous system were reviewed by the EPA and aspects of their strengths and limitations appropriately noted. The EPA's comments will not be repeated except where needed to place the EPA's comments with a larger context. The EPA correctly identified the key studies. Of these the Chen et al. (Chen et al., 2012) is the strongest. These authors mated SD rats in-house and culled litters to 8 (4 M and 4 F), randomized pups shortly after birth to create new mixed litters then rotated newly formed litters among the dams with the goal of distributing and neutralizing litter effects. They also used 40 litters in the experiment with 10 males and 10 females from different litters for testing, generating 320 offspring for assessment (4 M and 4 F per litter x 40 litters = 320). Progeny were treated with 0, 0.02, 0.2, or 2 mg/kg BaP by gavage on P5-11 with 1 M/F pair within each litter receiving one of the 4 doses. They divided the 320 offspring into 4 tracks or groups for testing. Each track contained 80 rats. Track-1 rats received two early tests (surface righting and cliff aversion) and open-field at P18. Track-2 rats received two different early tests (inclined plane/negative geotaxis and forelimb grip strength) and open-field at P20. Track-3 and Track-4 rats received postweaning testing. Track-3 received testing during adolescence: P34 open-field, P35 elevated plus maze (EPM), and P36-39 Morris water maze (MWM) whereas Track-4 rats received testing as adults: P69 open-field, P70 EPM, and P71-74 MWM. Most of the behavioral tests were standard but the MWM requires mention. The pool was 130 cm in diameter with a 9

cm platform. On day-1 rats were given a 60 s trial with no platform as acclimation; no measurements were taken. Spatial learning assessment occurred on the following 4 days with 4 trials/day with an ITI of 5 min. On day-5 rats received a 60 s probe trial with the platform removed after the last learning trial. Small but significant body weight reductions were seen on P36 and P71 in the 2 mg/kg group but none of the physical landmarks of development were affected. There were delays in surface righting in the low, mid and high dose groups but on different days; there were delays in the inclined plane/negative geotaxis test at all doses on P12 but only in the high dose on P14. In the open-field there were increases in activity and rearing at P34 and 69 but not at P18 or P20; these effects were mostly in the high dose group with one effect at P69 in the mid dose group on activity but not on rearing. There were also effects in the EPM at P70 with increased time in open, reduced latency to first open entry, increased number of open arm entries, and decreased entries into closed arms; these effects were seen in the 0.2 and 2 mg/kg BaP groups for latency and time in open at P70 but not at P35. On the third measure, number of open arm entries, there were increases at P70 in the 2 mg/kg males and in the 0.2 and 2 mg/kg group females. The fourth measure was number of closed arm entries. The sum of open and closed arm entries can be construed as a measure of activity. In this study the sum of these showed no group differences, therefore, this was suggested by the EPA to indicate that the time in open and open arm entries was a specific anxiety-related effect which in this case since the BaP groups showed higher levels this was interpreted as lower than normal anxiety. The effects were more prominent in the high dose group than in the mid dose group with no effects in the low dose group on the EPM test. But by far the most striking finding in this study was in the MWM. In both males and females, at both P36-39 and P71-74, high and mid dose BaP groups had longer escape latencies to find the hidden platform compared with Controls or the low dose group. On the probe trial, in both males and females at P40 and P75 time in the target quadrant and number of site crossovers were significantly reduced in the high dose group at both ages, and in adults also in the mid dose group and although no data are shown, it is stated that there were no swim speed differences on the probe trial.

Strengths: This study has a number of strengths; these included the care to use in-house breeding, using 40 litters, standardizing litter size, balancing for sex, testing multiple dose levels of BaP, administering BaP by gavage rather than by i.p injection, efforts to neutralize litter effects, use of multiple behavioral tests, appropriate ANOVA as the main way of analyzing the data (but see caveat below on post hoc testing method), and use of the MWM (but see below about issues with the MWM), use of a split-litter design.

Weaknesses: Despite these strengths, the study has weaknesses. The size of the MWM, while appropriate for the P36-39 rats, was undersized for adult rats. Mitigating this is the fact that BaP-related effects were seen despite the size of the maze. Another concern is the reliance on latency as the sole index of performance under the presumption that it accurately reflects learning when it may not. This is an issue of concern inasmuch as the mid and high dose groups were affected in parallel including on day-1 (at both ages). No sub-analysis by trial on day-1 was performed to determine if the groups started out equal or unequal. In addition, no cued trials were given to rule-out visual problems. No measurements of path length or swim speed were recorded on learning trials to rule-out other performance factors, and the probe trial was given immediately after the last learning trial thereby limiting its interpretive value. Mitigating some concern over

swim speed was the fact that the authors report (but do not show data) that there were no swim speed differences on the probe trials. Why the authors measured swim speed only on the probe trial and not on the learning trials is unclear. The use of the LSD a posteriori test is another concern as this test is inappropriate in a study with 4 groups. This test over-calls significant differences when there are more than three groups, as there were in this experiment, because it does not control for multiple comparisons. Had an appropriate a posteriori pairwise comparison been used, such as the Hochberg, or the False Discovery Rate, some of the smaller group differences reported as significant might not have been by other methods. In the EPA review of this study, the parallelism of the learning curves was noted. It was noted that this reflected equal learning in all groups, which is the correct interpretation. Only if we knew the individual performances on the 4 trials on day-1 could it be determined if the groups began the test equally or if the BaP groups were different from the outset. The EPA review also expressed concern about the interpretative value of the probe trial data in light of the fact that the affected BaP groups never reached the same level of proficiency on the learning trials as Controls, suggesting that they had not learned the platform's location sufficiently to be able to remember it as well as Controls on the probe trial. This too is a correctly identified concern. And there are other concerns. The pup randomization and litter rotation used frequently during the preweaning period is an unproven method of trying to prevent litter effects, but its effects are unknown. It may work as intended or it may introduce stress effects. While stress effects would be expected to be randomly distributed across litters, there exists the potential for BaP x stress interactions that this design might cause but would be impossible to detect using this design. Unfortunately, there are no studies in the literature that have tested this rotation design against a standard intact litter design to determine if it is an improvement over previous designs or a detriment. Another concern of unknown significance is of having all dose groups in the same litter. Potentially this could cause cross contamination of BaP from the higher dose groups to the lower dose groups or controls. Similarly, it is unknown if the dams could distinguish difference about the differently dosed pups and thereby differentially care for pups in different ways. If this were to happen it could introduce another unknown factor into the outcomes. Despite the above concerns and despite interpretational issues concerning whether the data reflect a spatial learning deficit, the MWM show a clear BaP-dependent effect that cannot be ignored. Rather than placing reliance on the EPM data and dismissing the MWM data, the committee recommends taking them and the other neurobehavioral data in this study collectively and viewing them in their totality as evidence of a developmental neurobehavioral effect of neonatal BaP exposure with long-term adverse CNS effects. While differential maternal responses to differently dosed pups and stress effects arising from litter rotation have indeterminate effects, it is going beyond the data to suggest that they undermine the study since they could just as easily turn out to be of no consequence for the developmental neurotoxicity of BaP. This is important since other studies provide support to the findings of Chen et al. (2012).

Li et al. (Li et al., 2012) conducted an experiment using an inbred mouse strain with a Loss of Function (LOF) mutation in the Cpr gene which encodes for the P450 enzyme oxidoreductase which is involved in BaP metabolism. This is a specialized experiment to test a specific hypothesis. It is of interest because the KO and WT mice were given BaP on E14-17. BaP was administered by inhalation at a dose of 100 g/m³. Of particular interest in terms of developmental neurotoxicity was that among other parameters assessed in the offspring, mice

were tested on an object discrimination task which was a modified version of the better known Novel Object Recognition test (NOR). Setting the details aside, the upshot was that the BaP-exposed KO mice, but not BaP-exposed WT mice, showed a marked reduction in novel object preference suggesting a hippocampally-mediated non-spatial learning deficit. Because the effect occurred only in the KO mice that were deficient in metabolizing BaP, the data suggest that BaP is more toxic in those with reduced oxidoreductase capacity. In humans this could occur by interindividual CNV or SNP differences causing some people to be more susceptible to BaP than others. Unfortunately, only 4-5 mice were tested per group in a test known for its variability. This reduces confidence that the effect is valid.

Bouayed et al. (Bouayed et al., 2009) also used mice. In this experiment Swiss albino mice were treated with 0, 2 or 20 mg/kg BaP by gavage on P0-14 and developmental parameters and behavior assessed at different ages. Assessments included physical development, maternal behavior (nest building and pup retrieval), surface righting, inclined plane (a.k.a. negative geotaxis), forelimb grip suspension; open-field on P15, water escape pole climbing on P20, EPM on P32, and spontaneous alternation on P40. No effects of BaP were found on physical development or maternal behavior. Delays in surface righting were found in both BaP groups on P3 and 5, on inclined plane in the high dose group on P5, 7, and 9, on the wire suspension test on P9 and 11, with no effects in the open-field, delays in males in the high dose group on the water escape test, and increased time in open and related measures in the EPM. One low dose effect was also seen as increased alternation frequency in the Y-maze, an effect not seen at the dose 10 times higher. **Strengths:** This is one of the few developmental neurotoxicity experiments in mice and therefore provides some species convergent data of developmental neurotoxicity. The study also included testing more than one dose of BaP, multiple behavioral tests, and appropriate statistical analyses. **Weaknesses:** Only 5 litters were used in each group and there is no evidence that litter effects were accounted for. Many of the tests, while affected, are of limited interpretative value because they may represent transient delays from which full recover may occur, and the 20 mg/kg dose of BaP used was too high to be very relevant.

Tang et al. (Tang et al., 2011) treated Wistar rats starting at weaning for 14 weeks with 1, 2.5, or 6.25 mg/kg BaP i.p. from approximately P21 past 200 days of age and assessed the animals in the Morris water maze (MWM) to a hidden platform as a test of spatial learning starting one day after the end of treatment. In this procedure rats were tested in a circular pool 180 cm in diameter and apparently given 1 trial/day although the authors do not specify this parameter and it may have been several trials per day. They found significant increases in maze latency on all 5 days of testing in the 2.5 and 6.25 mg/kg BaP dose groups but only on day-3 in the 1 mg/kg dose group. They gave a reference memory (probe) trial after the last learning trial on day-5. On this trial, they found effects of BaP at all doses on platform site crossovers and they found reductions in target quadrant bias in the 2.5 and 6.25 mg/kg BaP dose groups. **Strengths:** They tested multiple doses, groups sizes (9/group) were minimally adequate, the maze was appropriately sized for rats, reasonable learning curves were obtained, and the data appropriately analyzed. **Weaknesses:** Latency is a potentially problematic index of learning because it can be affected by performance factors, such as swim speed, an issue the authors fail to address. Also, the probe trial was given shortly after the last learning trial therefore it cannot be determined if the effects were on working or reference memory. Also the probe trial was too long at 120 s; it is known

that spatial bias progressively deteriorates after 30 s. This is mitigated by the fact that the effects of BaP were significant even with a long probe trial. More importantly, while treatment began on approximately P21, this was not an early but rather a late developmental exposure period that extended well into adulthood, including during testing. Therefore, it is not clear that the effects were irreversible since testing commenced shortly after the last treatment rather than allowing for a no-treatment period to intervene between the end of treatment and testing in order to determine the permanence of the effects.

Qiu et al. (Qiu et al., 2011), similarly to Tang et al. (2011) above, gave Sprague-Dawley male rats 6.25 mg/kg BaP i.p. but in this study they started at P28 and treated the rats for 14 weeks. Rats were tested an unspecified number of days after the last treatment in a small 130 cm diameter MWM with a 9 cm hidden platform. They gave 4 trials/day from different start locations for 5 days following a habituation day in the pool with no platform present as acclimation. Apparently the probe trial was given on the last day of platform training. They found a significant increase in latency to find the platform across all 5 days of testing and a reduction in the number of platform site crossovers and time spent in the target quadrant on the probe trial.

Strengths: They used 8 rats/group, a minimally sufficient sample size, the data were appropriately analyzed, and the MWM procedures were generally appropriate (with some caveats).

Weaknesses: A 130 cm maze for adult male SD rats is too small to provide a good test of spatial navigation. Adult rats should be assessed in mazes no less than 183 cm (6 ft.) in diameter. Probe trials should be given 24 h or more hours after the last learning trial, and latency is a potentially confounded index of learning and should be cross-validated against swim speed and/or analysis of path length, neither of which were reported in this experiment. But the greatest concern about this study is that the BaP and Control groups differed significantly on Day-1 of MWM testing. This raises the concern that the BaP animals started out the test performing differently. It is a fundamental concept in learning and memory that if groups start out different and learn in parallel to control they are likely to be different because of a performance difference unrelated to learning. This can be resolved by examining the trials on day-1 individually. Ideally, both groups start out the same on trial-1 when none of the animals know where to go to find the platform. If the groups begin to diverge on trials after the first or second one it suggests that the treated animals are less able to find or remember where the platform compared with controls. In such cases one has to consider whether the treated animals have impaired swimming ability or vision and therefore have secondary sensorimotor impairments that reduce their ability to perform the spatial aspects of the task. Unfortunately, the authors did not address the issue thereby leaving it unresolved. This experiment is also not a test of early, but rather of late, developmental effects yet the data are consistent with those of Chen et al. (2012).

Xia et al. (Xia et al., 2011) like Qiu et al. (2011) used male SD rats and started treatment at P28 and treated for 13 weeks. They used 8 rats/group and the dose groups were Control, 1, 2.5, and 6.25 mg/kg BaP given daily by i.p. injection dissolved in DMSO then diluted with corn oil. In

this experiment, rats were tested in the MWM before BaP treatment (where no group differences were found) and after the end of treatment. This maze was also 130 cm in diameter and platform size was unspecified. For the post-treatment MWM assessment, rats were given 4 trials per day for 5 days with a probe trial given shortly after the last learning trial on day-5. Significant

increases in escape latencies were found in the 2.5 and 6.25 mg/kg BaP groups and as in the Qiu et al. study, the effects were uniform on all days including day-1, again raising concern about swim speed or other interfering performance effects of the compound such that the animals in the treated groups may not have started the test equally capable of performing it. On the probe trial, an effect of 6.25 mg/kg BaP dose was found on platform site crossovers and on time in the target quadrant. Standard control methods to rule-out possible sensorimotor deficits would be to conduct cued trials with a visible platform with curtains closed around the maze to prevent use of distal cues, to track swim speed during learning trials, and to report path length, which is largely immune from speed effects. Strengths: The study has minimally sufficient sample sizes, it included 3 BaP dose levels and two controls groups (vehicle and what they refer to as 0 mg/kg), the data were appropriately analyzed, and the effects at the two higher doses were clear-cut. Weaknesses: As in several of the studies, concerns exist about the small size of the maze for adult male rats, the reliance on latency without convergent measures less prone to confounding, the differences on day-1 of the test with no analysis of day-1 data trial-by-trial, and the fact that the probe trial was not given 24 h or more after the last learning trial.

In a study by Maciel et al. (Maciel et al., 2014) motor and cognitive effects were assessed in Wistar rats. However, this study's relevance to the current assessment is marginal since the exposure was in adult rats, but the data nevertheless support the view that BaP is neurotoxic in adult as well as developing animals.

More relevant is a study by Patri et al. (Patri et al., 2013). In this unusual design, P5 Wistar rats were given a single intracisternal injection of 0.1 μ M of BaP. The rats were raised and tested in a MWM before 6 weeks of age. Starting on P28, rats were tested in a 143 cm diameter maze for 8 days, 4 trials/day with a probe trial given 24 h after the last learning trial. The BaP group had significantly longer escape latencies than untreated or vehicle treated controls on days 3-8. Significantly, not only were the treated group's latencies longer, they had much longer path lengths than controls. Furthermore, swim speed was assessed and no differences found. On the probe trial, the BaP groups has fewer site crossovers and reduced time in the target quadrant. Strengths: This experiment conducted the MWM better than in any of the above studies because they appropriately accounted for and eliminated concerns over potential swim speed differences by directly measuring swim speed and analyzing path length. They also showed that the groups began the test with essentially identical performance. They also conducted the probe trial 24 h after the last learning trial, making a reference memory deficit apparent without confounding with possible working memory effects. Weaknesses: The intracisternal route of BaP administration makes this study difficult to utilize to compare to anything else. In addition, the groups sizes were marginal: N = 4 in the untreated group, N = 7 in the DMSO-vehicle group, and N = 8 in the BaP group. In addition, it is not stated how many litters the rats came from leaving concern that they may have been drawn from a small number of litters without attention to

proper litter sampling. Despite these limitations, the data further support that BaP is developmentally neurotoxic.

Mechanistic Studies: There are studies implicating plausible biological modes of action of BaP on brain development. Brown et al. and McCallister et al. gave gravid LE rats 25 or 150 (Brown et al., 2007) or 300 mg/kg (McCallister et al., 2008) BaP on E14-17 and found metabolites in higher concentrations in brain than liver and that BaP reduced mRNA of NMDA-NR2A and NR2B and AMPA receptor expression and protein concentrations in hippocampus and inhibited NMDA-dependent cortical barrel field post-stimulation spikes by 50%. Bouayed et al. gave Swiss mice 2 or 20 mg/kg by gavage on P0-14 and found 2 mg/kg effects on surface righting, forelimb grip, EPM similar to that found by Chen et al., reduced spontaneous alternation, and reduced brain mRNA expression of 5-HT1A receptor (Bouayed et al., 2009). The quality of some of the studies was limited. For example, in Bouayed et al (2009) and McCallister et al (2008), there were insufficient number of litters, litter effects were not accounted for and/or subjective behaviors were not evaluated blind to treatment group. These and other quality issues that were not identified in the EPA report. Nevertheless, these and related studies implicate NMDA and AMPA glutamate receptors, as well as 5-HT receptors as potentially mediating the neurobehavioral effects seen by Chen et al. (2012) and others and support the view that developmental exposure to BaP adversely effects brain development and behavior.

Synthesis: The above neurodevelopmental studies provide evidence that BaP induces developmental neurotoxicity in animals and humans. Several of the studies were sufficiently well done that when reading across the human, animal, and mechanistic studies, they provide evidence of developmental neurotoxicity and they benefit from convergent data done in adolescent and adult rodents with BaP exposure and finding neurobehavioral toxicity thereby buttressing the prenatal and neonatal data. Nevertheless, each of the studies has limitations. This applies to tests known to show experiment-to-experiment and cross-laboratory variability. These include the elevated-plus and novel object recognition tests. Studies using these methods should be replicated, ideally by another lab, where similar effects are found before the evidence would rise to the level of "strong." There are many examples in the literature where findings with these tests cannot be replicated. Methods such as the open-field test of locomotor activity are more reliable provided the test is properly done. This includes using an automated system, testing for a sufficient length of time (30-60 min, rather than 5 min as in Chen et al.), and proper environmental controls. The MWM used in a number of BaP animal studies, largely in the absence of other tests of learning and memory. While the MWM is a superb test when properly conducted to assess spatial learning and reference memory, and is a strongly hippocampally-dependent form of cognition, it is the case that the above datasets do not have the benefit of convergence by having other tests of learning and memory to cross-validate the MWM findings. Conversely, the fact that there are multiple experiments using the MWM increases the confidence that developmental BaP has reliable effects on MWM performance, and this is a strength of this set of experiments taken as a whole. Clearly there are deficiencies in the MWM methods in every experiment reviewed. This raises concern about how much weight should be placed on these data. As noted, failure to include proper maze scaling, concerns over not including control procedures for non-cognitive performance factors, and the learning curves being parallel in the Chen et al. and other studies with later BaP exposures raises concern about

whether these are true spatial learning deficits or performance effects. In the final analysis the weight of evidence for developmental neurotoxicity using a read-across approach of the three categories of evidence (human, animal, mechanistic) the committee concluded that BaP is a neurodevelopmental neurotoxic agent.

2b. Reproductive toxicity (sections 1.1.2, 1.2.1). The draft assessment concludes that male and female reproductive effects are a human hazard of benzo[a]pyrene exposure. Do the available human and animal studies support this conclusion?

2c. Immunotoxicity (sections 1.1.3, 1.2.1). The draft assessment concludes that immunotoxicity is a potential human hazard of benzo[a]pyrene exposure. Do the available human and animal studies support this conclusion?

2d. 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?

RESPONSE

There are other types of non-cancer toxicity associated with BaP and the EPA review identifies these. For nervous system, the most notable of these, as reviewed in the EPA report, are the BaP animal experiments where BaP was administered starting at weaning, adolescence, or to adult rodents. The committee concurred with the EPA that these represent additional types of non-cancer BaP toxicity. The only difference is that the committee suggested that the EPA include these in its overall assessment of BaP as both a developmental and adult neurotoxic agent, as it was not clear in the report what the cutoff was for placing a study in the developmental versus non-developmental category given that there are prenatal, neonatal, weaning, and adolescent exposure studies, all of which are developmental in one sense or another even apart from the adult neurotoxicity exposure studies. The EPA report clearly included the prenatal and early postnatal studies in the developmental neurotoxicity section, but placed the weaning (starting exposure at P21) and adolescent (starting exposure at P28) in the "other" non-cancer nervous system section. Further justification of the boundaries would be useful.

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.

3a. Oral reference dose for effects other than cancer (section 2.1). The draft assessment proposes an overall reference dose of 3×10^{-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 implicit for exposures during a critical window of development?

Developmental Toxicity: The draft assessment concludes that developmental toxicity and developmental neurotoxicity are human hazards of BaP exposure. Do the available human, animal, and mechanistic studies support this conclusion?

RESPONSE:

Although the human data by Perera et al. and Tang et al. are from exposures to PAH mixtures and rely on evidence from BaP DNA-adducts measured in maternal and cord blood, these data from the Columbia and Tongliang child development cohort studies find consistent evidence of long-term effects from preschool up through at least age 9 so far (for the most recently reported data from the Columbia) on outcomes such as the ADHD impulsive-inattentive and other neuropsychological outcomes as well as the early findings of effects on birth weight, length, and head circumference.

In animals, the data of Chen et al. (2012) with 40 litters and 320 offspring tested using control and 3 dose levels of BaP given orally on P5-11 are important data for dose-response determinations. With findings on multiple behaviors including most significantly on the open-field, EPM and MWM along with data from several other studies, including data in adolescent and adult animals finding neurobehavioral effects at doses similar to those used by Chen et al., support the conclusion that BaP is neurotoxic to the developing nervous system both early and later. Although methodological concerns were identified in the studies, the overall pattern of effects and convergence across studies support the view reached by the EPA that the evidence is sufficient to use for dose-response analysis. In terms of mechanism or mode of action on the nervous system there is no one or even predominant line of evidence supporting a specific molecular pathway for BaP. There are multiple studies implicating electrophysiological changes, effects on monoamine neurotransmitters and their receptors, including changes in mRNA expression of receptors particularly glutamatergic and 5-HT1A receptors. While there is no direct link between these and the neurobehavioral findings, they provide biologically plausible evidence for the behavioral effects. Collectively, the data support the use of the Chen et al. data as the key evidence for conducting a dose-response analysis. In the EPA report supplement the agency uses the BMD method. The committee was in general agreement with this, but rather than use the EPM data, the committee found that the MWM data, including on both the learning and probe trials on this test were more compelling and should be used instead.

The EPA report states that with regard to the Chen et al. (2012) study for the EPM finding:

“These results indicate effects on a single, discrete neurological function that are unlikely to be complicated by changes in other processes such as motor activity (total activity, calculated by summing open and closed arm entries was unchanged with treatment). This neurobehavioral endpoint is supported by similar observations in developing (Bouayed et al., 2009a) and adult (Grova et al., 2008) mice, and may be indirectly related to

observations of increased aggression in mice (Bouayed et al., 2009b) and is considered adverse.”

But the committee noted that the EPM has a number of limitation of greater importance than those associated with the MWM. The EPM can only be used once to sample behavior whereas the MWM samples behavior over many trials over many days. The EPM is conducted for approximately 5 min because once the novelty wears off, it no longer induces the conflict between open and closed spaces to create an approach-avoidance conflict. The MWM shows a learning curve over many trials. The EPM is sensitive to many factors such as the animal's handling history, testing conditions and others. While the test is widely used, it has also been criticized. The test remains in use mostly to because of a lack of alternatives not because the test is robust. Alternative methods, including the elevated platform test, have been proposed recently but are not yet in wide use. Rather than place so much emphasis on the EPM findings, the committee recommends using the EPM and MWM data together and interpret them collectively. Since the lowest BMD of the several outcomes across these tests is obtained with the MWM, the committee recommends using the lowest BMD from this test to provide the best basis for determining a POD for adding UFs.

Reference List

- Bouayed J, Desor F, Rammal H, Kiemer AK, Tybl E, Schroeder H, Rychen G, Soulimani R (2009) Effects of lactational exposure to benzo[alpha]pyrene (B[alpha]P) on postnatal neurodevelopment, neuronal receptor gene expression and behaviour in mice. *Toxicology* 259:97-106.
- Brown LA, Khoubouei H, Goodwin JS, Irvin-Wilson CV, Ramesh A, Sheng L, McCallister MM, Jiang GC, Aschner M, Hood DB (2007) Down-regulation of early ionotropic glutamate receptor subunit developmental expression as a mechanism for observed plasticity deficits following gestational exposure to benzo(a)pyrene. *Neurotoxicology* 28:965-978.
- Chen C, Tang Y, Jiang X, Qi Y, Cheng S, Qiu C, Peng B, Tu B (2012) Early postnatal benzo(a)pyrene exposure in Sprague-Dawley rats causes persistent neurobehavioral impairments that emerge postnatally and continue into adolescence and adulthood. *Toxicol Sci* 125:248-261.
- Li Z, Chadalapaka G, Ramesh A, Khoshbouei H, Maguire M, Safe S, Rhoades RE, Clark R, Jules G, McCallister M, Aschner M, Hood DB (2012) PAH particles perturb prenatal processes and phenotypes: protection from deficits in object discrimination afforded by dampening of brain oxidoreductase following in utero exposure to inhaled benzo(a)pyrene. *Toxicol Sci* 125:233-247.
- Maciel ES, Biasibetti R, Costa AP, Lunardi P, Schunck RV, Becker GC, Arbo MD, Dallegrave E, Goncalves CA, Saldiva PH, Garcia SC, Leal RB, Leal MB (2014) Subchronic oral administration of Benzo[a]pyrene impairs motor and cognitive behavior and modulates S100B levels and MAPKs in rats. *Neurochem Res* 39:731-740.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

McCallister MM, Maguire M, Ramesh A, Aimin Q, Liu S, Khoshbouei H, Aschner M, Ebner FF, Hood DB (2008) Prenatal exposure to benzo(a)pyrene impairs later-life cortical neuronal function. *Neurotoxicology* 29:846-854.

Patri M, Singh A, Mallick BN (2013) Protective role of noradrenaline in benzo[a]pyrene-induced learning impairment in developing rat. *J Neurosci Res* 91:1450-1462.

Perera F, Li TY, Lin C, Tang D (2012a) Effects of prenatal polycyclic aromatic hydrocarbon exposure and environmental tobacco smoke on child IQ in a Chinese cohort. *Environ Res* 114:40-46.

Perera FP, Chang HW, Tang D, Roen EL, Herbstman J, Margolis A, Huang TJ, Miller RL, Wang S, Rauh V (2014) Early-life exposure to polycyclic aromatic hydrocarbons and ADHD behavior problems. *PLoS ONE* 9:e111670.

Perera FP, Li Z, Whyatt R, Hoepner L, Wang S, Camann D, Rauh V (2009) Prenatal airborne polycyclic aromatic hydrocarbon exposure and child IQ at age 5 years. *Pediatrics* 124:e195-e202.

Perera FP, Rauh V, Whyatt RM, Tang D, Tsai WY, Bernert JT, Tu YH, Andrews H, Barr DB, Camann DE, Diaz D, Dietrich J, Reyes A, Kinney PL (2005) A summary of recent findings on birth outcomes and developmental effects of prenatal ETS, PAH, and pesticide exposures. *Neurotoxicology* 26:573-587.

Perera FP, Rauh V, Whyatt RM, Tsai WY, Bernert JT, Tu YH, Andrews H, Ramirez J, Qu L, Tang D (2004) Molecular evidence of an interaction between prenatal environmental exposures and birth outcomes in a multiethnic population. *Environ Health Perspect* 112:626-630.

Perera FP, Tang D, Wang S, Vishnevetsky J, Zhang B, Diaz D, Camann D, Rauh V (2012b) Prenatal polycyclic aromatic hydrocarbon (PAH) exposure and child behavior at age 6-7 years. *Environ Health Perspect* 120:921-926.

Perera FP, Wang S, Vishnevetsky J, Zhang B, Cole KJ, Tang D, Rauh V, Phillips DH (2011) Polycyclic aromatic hydrocarbons-aromatic DNA adducts in cord blood and behavior scores in New York city children. *Environ Health Perspect* 119:1176-1181.

Qiu C, Cheng S, Xia Y, Peng B, Tang Q, Tu B (2011) Effects of subchronic benzo(a)pyrene exposure on neurotransmitter receptor gene expression in the rat hippocampus related with spatial learning and memory change. *Toxicology* 289:83-90.

Tang D, Li TY, Liu JJ, Chen YH, Qu L, Perera F (2006) PAH-DNA adducts in cord blood and fetal and child development in a Chinese cohort. *Environ Health Perspect* 114:1297-1300.

Tang D, Li TY, Liu JJ, Zhou ZJ, Yuan T, Chen YH, Rauh VA, Xie J, Perera F (2008) Effects of prenatal exposure to coal-burning pollutants on children's development in China. *Environ Health Perspect* 116:674-679.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Tang Q, Xia Y, Cheng S, Tu B (2011) Modulation of behavior and glutamate receptor mRNA expression in rats after sub-chronic administration of benzo(a)pyrene. *Biomed Environ Sci* 24:408-414.

Xia Y, Cheng S, He J, Liu X, Tang Y, Yuan H, He L, Lu T, Tu B, Wang Y (2011) Effects of subchronic exposure to benzo[a]pyrene (B[a]P) on learning and memory, and neurotransmitters in male Sprague-Dawley rat. *Neurotoxicology* 32:188-198.

Dr. Christi Walter

Response to Charge Question 1. Literature Search/Study Selection and Evaluation

The processes for searching, selecting and evaluating literature were clearly described. The most recent references are around 2011. References for reproductive toxicology have been provided in the comments for that section.

Response to Charge Question 2b. Reproductive Toxicity

Section 1.1.2

The current document conveys a summary of traditional toxicological outcomes of BaP on reproduction. Because spermatogenesis is an ongoing process that renews itself, it is important to distinguish between an immediate effect and a lasting effect on male germ cells. In contrast, since oocytes develop in utero and are not continuously renewed, it is likely to be a major difference between male and female germlines. The distinction between immediate and long lasting effects on reproduction is rarely made in the literature so it is not surprising that it is not covered in the current document. There are many instances when an agent disrupts spermatogenesis, but the short term effects can be lost and long term normal spermatogenesis restored. This is an important aspect of male reproductive biology. If the intent for this review is to advise only about the immediate effects of BaP exposure on male reproduction, then the document is largely adequate. However, if the full ramifications of BaP exposure on reproduction are intended, there should be some discussion of the timeframe between treatment and observations and whether or not there was time for an additional wave or more of spermatogenesis before the outcomes were measured. Have the measurements been performed before spermatogenesis is restored, or after? The reversal of short term effects can involve testis size and weight, since this is a direct reflection of the amount of spermatogenesis, number and shape of sperm in the epididymis, count of ejaculated sperm, and histological appearance of the testis. However, it is also true that high concentrations of toxic agents may kill spermatogonial stem cells and have more permanent effects of spermatogenesis, testis, size and weight, and etc. Distinguishing between the different outcomes is important if we seek to inform that there can be short term consequences, but with enough time, many/most will go away.

An aspect of the document that would benefit from additional consideration is life stage and cell type. Because these cells will direct development of the next generation, successful reproduction may be compromised if germ cell mutagenesis is increased. *De novo* germ line mutations can result in genetic disease, miscarriage, infertility, etc. Life stage at exposure is of critical importance. Pre-spermatogonial stem cells proliferate extensively while migrating to and colonizing the embryonic gonad and after birth. This is a window of susceptibility to mutagenesis that could result in lifelong increased mutant frequency in mature germ cells that would subsequently affect reproductive outcome (Xu et al., 2014). Further, stem cells are on the unprotected side of the testis blood barrier and more likely to be exposed to higher amounts of genotoxins. Because the stem cells are the cells that will continue to give rise to sperm, the impact

of mutagenesis on stem cells can affect reproduction. There are no direct studies of the effects of BaP on spermatogonial stem cell mutagenesis, but there is a reference that implicates stem cell mutagenesis (Olsen et al., 2010). There are additional papers on the effects of BaP on adduct formation, mutagenesis, and gene expression (Verhofstad et al., 2010a; Verhofstad et al., 2010b; Verhofstad et al., 2011). To the best of my knowledge no studies on the mutagenic effects on oocytes has been performed and is likely due to the difficulty in obtaining adequate numbers of cells.

There are few studies on BaP effects on ovary, oocytes, etc. I am providing 4 I did not find in the HERO database, to help shore up the interpretations.(Einaudi et al., 2014; Kummer et al., 2013; Sadeu and Foster, 2011; Sadeu and Foster, 2013).

The tables figures to be particularly helpful. However, the document contains far too many abbreviations if the intention is to make the document understood by non-experts. While many who will read the report will be familiar with the abbreviations, it is dismaying if you aren't familiar with them.

Available studies do support the conclusion that BaP exerts hazardous effects on human reproduction.

Recommended additional references for reproductive toxicology:

Einaudi, L., B. Courbiere, V. Tassistro, C. Prevot, I. Sari-Minodier, T. Orsiere, and J. Perrin. 2014. In vivo exposure to benzo(a) pyrene induces significant DNA damage in mouse oocytes and cumulus cells. *Human Reproduction*. 29:548-554.

Kummer, V., J. Maskova, Z. Zraly, and M. Faldyna. 2013. Ovarian disorders in immature rats after postnatal exposure to environmental polycyclic aromatic hydrocarbons. *Journal of Applied Toxicology*. 33:90-99.

Olsen, A.-K., Å. Andreassen, R. Singh, R. Wiger, N. Duale, P.B. Farmer, and G. Brunborg. 2010. Environmental exposure of the mouse germ line: DNA adducts in spermatozoa and formation of *de novo* mutations during spermatogenesis. *PLoS ONE*. 5:e11349.

Sadeu, J.C., and W.G. Foster. 2011. Effect of in vitro exposure to benzo a pyrene, a component of cigarette smoke, on folliculogenesis, steroidogenesis and oocyte nuclear maturation. *Reproductive Toxicology*. 31:402-408.

Sadeu, J.C., and W.G. Foster. 2013. The cigarette smoke constituent benzo a pyrene disrupts metabolic enzyme, and apoptosis pathway member gene expression in ovarian follicles. *Reproductive Toxicology*. 40:52-59.

Verhofstad, N., J. La Pennings, C.T.M. van Oostrom, J. van Benthem, F.J. van Schooten, H. van Steeg, and R.W.L. Godschalk. 2010a. Benzo(a)pyrene induces similar gene expression changes in testis of DNA repair proficient and deficient mice. *Bmc Genomics*. 11.

5/4/15 Revised Comments for review and deliberations by the CAAC Committee Augmented for the Review of EPA's Draft IRIS Benzo[a]pyrene Assessment. Do Not Cite or Quote. These comments are draft and work in progress. They do not reflect consensus advice or recommendations, have not been reviewed or approved by the chartered SAB and do not represent EPA policy.

Verhofstad, N., C.T.M. van Oostrom, J. van Benthem, F.J. van Schooten, H. van Steeg, and R.W.L. Godschalk. 2010b. DNA Adduct Kinetics in Reproductive Tissues of DNA Repair Proficient and Deficient Male Mice After Oral Exposure to Benzo(a)pyrene. *Environmental and Molecular Mutagenesis*. 51:123-129.

Verhofstad, N., C.T.M. van Oostrom, E. Zwart, L.M. Maas, J. van Benthem, F.J. van Schooten, H. van Steeg, and R.W.L. Godschalk. 2011. Evaluation of Benzo(a)pyrene-Induced Gene Mutations in Male Germ Cells. *Toxicological Sciences*. 119:218-223.

Xu, G., C.A. McMahan, and C.A. Walter. 2014. Early-Life exposure to benzo[a]pyrene increases mutant frequency in spermatogenic cells in adulthood. *PLoS ONE*. 9:e87439.

Inhalation Reference Concentration

Section 2.3.1-5

Inhalation is a common route of exposure, thus it is important to have an RfC since BaP is a near-ubiquitous pollutant. While the RfC for inhalation of BaP is important, the database to support an RfC is thin. There are many assumptions, manipulation through conversion factors and extensive extrapolation of the limited data. Unfortunately, the limited number of studies and inadequacies of published studies to address the criteria needed for establishing a robust RfC, result in a very low confidence in the RfC and lead to questions as to whether it is worth publishing. Additional comments are provided below.

Page 2-16 lines 10-11. Human inhalation data, which had been discounted as useful for determining the RfC on page 2-15 is cited as supporting the animal data. Discounted data should not be used to support animal data. Rather, they are consistent with the animal data, which is also inadequate for determining an RfC.

Not clear how results from a different route of exposure can be considered to bolster the inhalation effects when routes are considered separately? Page 2-23 lines 33-34.