

Draft Benzo[a]pyrene Charge Question Responses

Charge Question #1: Literature search and study selection

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

- **Literature Search**

- Well documented and clearly described. Figure LS-1 very helpful
- Suggested Clarifications/Enhancements
 - Review of references within the primary and secondary literature are also used to identify potentially relevant publications.
 - “secondary” literature searches – to avoid potential bias in endpoints searched for. As evidence for additional effects (e.g. cardio) or specific data gaps (e.g., MOA) emerge a secondary search with additional search terms is conducted

- **Selection of Studies**

- Identification of Excluded Chemicals

- General exclusion criteria listed (e.g., inadequate reporting) are appropriate
- Suggests to increase transparency and clarity
 - Include a table containing the list of excluded references grouped by the applicable exclusion criteria in the supplementary information
 - Requirement of a direct measure of B[a]P was too restrictive for hazard identification. Epi studies of coke oven workers and other occupational groups with known exposures to B[a]P are valuable sources of information for determining causality even if they do not include quantification of B[a]P. These studies should at least be reviewed in the tables if not the text.
 - Some question exclusion of all animal study data regarding mixtures.

- Panel appreciates that EPA is developing a handbook outlining tools and processes to address study quality and risk of bias. In the interim EPA should provide sufficient detailed criteria for each step of the process leading to the selection of key studies for point of departure (POD) assessment. This will ensure that not only the rationale for initial study inclusion or exclusion are clearly understood, but also the strengths and weakness of studies selected (as well as those that are not) for POD assessment are fully transparent. Suggest EPA consider identifying these criteria in one location within the Literature Search and Study Selection, rather than directing the reader to other sections/references.
- Identify Additional Peer-Reviewed Studies that should be considered
 - Additional studies were identified by Panel members. A list will be compiled.

2. Hazard Identification

2a. Developmental toxicity

Developmental toxicity and Developmental neurotoxicity

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?

- Human data: In general the human data support that BaP is a human developmental toxicant and developmental neurotoxicant.
- Although the human data come from studies of PAH mixtures and cannot be definitively attributed to BaP alone, the method used to assess BaP DNA adducts was specific for BaP and not a source of concern. These adducts were measured in maternal and umbilical cord blood and correlated with personal air monitor measurements. Children were followed from birth to 9 years of age and show evidence of compromised developmental quotients, increased Attention Deficit Hyperactivity Disorder (ADHD) impulsivity and inattention, increased anxiety and depression, reduced birth weight, length, and head circumference and interactions between DNA adducts and environmental tobacco smoke resulting in lower full scale and verbal IQ on the Wechsler Primary and Preschool Scale of Intelligence (WPPSI).

- Animal data-

- Developmental: Yes, at a high degree of confidence.
 - B[a]P exposure in utero has been demonstrated to cause fetal death, affect fetal germ cells, and is a teratogen (see Shum et al Teratology 20(3)365 1979).
- Neurobehavioral: Yes at a moderate level of confidence. Multiple studies were identified that showed developmental neurotoxic effects (including Bouayed et al. (2009) and Chen et al. (2012) (Fig. 1-2, p. 1-18) but the committee recommended taking all the developmental and neurodevelopmental studies into account collectively. The committee noted that with regard to the key study EPA focused on of Chen et al. (2012) :
 - **Strengths:** Tested 80 offspring, 10M/10F from 40 litters, control and 3 dose levels, assessed for reflex development, open-field, Elevated Plus Maze, and Morris Water Maze at 2 ages (40 at P35 and 40 at P70). Most tests were appropriately conducted. Both males and females were tested.
 - **Weaknesses:** Morris Water Maze swim speed was not measured on learning trials (but was on probe trials), cued control trials were not included, post hoc method (Least Significant Difference) test was not appropriate, litter randomization and pup rotation among dams was raised as a concern because of its unknown effects, learning curves were parallel rather than convergent. There was mild oversampling in the statistical analysis by combining males and females from the same litter in the analysis without accounting for litter as a factor.
 - **On balance,** strengths outweighed weaknesses: clear dose-response Elevated Plus Maze and Morris Water Maze effects were found at the 0.2 and 2 mg/kg doses, a few instances of effects at 0.02 mg/kg were reported (reflex developmental and Elevated Plus Maze), and the sample size provided reasonable power to detect effects, the effects are consistent with other studies.

- Mechanistic studies:

- Developmental: Limited but plausible.

- Multiple studies have shown that B[a]P affects rapidly dividing cells. EPA should consider if inclusion of known mechanisms of action of B[a]P (e.g. cell division, reactive oxygen species) which are directly applicable to developmental toxicity warrant inclusion/reference.

- Neurobehavioral: Limited but plausible.

- 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) or 300 mg/kg (McCallister) BaP on E14-17 and found metabolites in higher concentrations in brain than liver and that BaP reduced mRNA of the NMDA-NR2A and NR2B and AMPA glutamatergic 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, Elevated Plus Maze similar to that found by Chen et al., reduced spontaneous alternation, and reduced brain mRNA expression of the serotonin-1A receptor.
 - 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 and will be provided in the Panel's report.
 - These and other studies implicate NMDA and AMPA glutamate receptors, as well as and serotonin receptors as potentially mediating the neurobehavioral effects seen by Chen and others and support the view that developmental exposure to BaP adversely effects brain development and behavior.

Question 2B. Reproductive Toxicity

(McIntyre, Walter, Moorthy and Poirier)

- Agree with EPA that B[a]P is a male and female reproductive toxicant in rodents via oral or inhalation routes of exposure
 - Reproductive toxicity is supported by mode of action/mechanistic studies
 - Additional references to support provided in group write-up
- Although not definitive evidence of a causal relationship between B[a]P exposure and reproductive toxicity in humans, findings in humans exposed to PAHs are consistent with those observed in laboratory animals, indicating a likely contribution of B[a]P to the adverse response.

- Consideration of study selection/expansion of endpoints
 - Suggest that EPA explore identified literature (noted in group write up) on ovarian follicle counts and DNA/mutagenesis in the testes; and for appropriateness for POD/BMD analyses and RfD determination.
 - May provide a mechanistic and biologically driven approach for POD
 - EPA should provided context as to the applicability of the inflammatory cervical response described in the Gao study for BMD/RfD generation
 - EPA may want to consider if this finding should be categorized under “reproductive effect”- or “other toxicity”)

Summary Recommendations:

- Review timing of male exposure and effect (may explain consistency of response)
- Examine suggested literature for dose-response effects on ovarian follicle counts
 - Further support of mode of action
 - Appropriateness for POD/BMD analyses and RfD determination
- Consider impact of B[a]P genotoxic effects on germ cells with respect to increased DNA damage and mutagenesis
 - Potential/likely mechanism of action
 - Appropriateness for POD/BMD analyses and RfD determination
- Provide additional clarity as to why certain studies (or parts of studies) are subsequently brought forward, whereas others were not.
- The inflammatory response described in the Gao study should be considered for BMD/RfD generation.

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?

Burchiel, Choi, English

- Yes, we believe that the available immunotoxicity data based on animal models of pure benzo[a]pyrene (BaP) and complex mixture exposures to humans (coke oven workers) supports the claim that BaP is a human hazard for the immune system.
- In vitro human PBMC studies should be included that support an understanding of Mechanisms of Action; while there is no doubt that BaP and other PAHs with specific SARs can cause suppression in human HPBMC at low concentrations in vitro, it is unclear whether these levels of exposure can be achieved in vivo with environmental inhalation exposures or ingestion of cooked foods.
- Immunotoxicity is caused by a combination of genotoxicity (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; compromising the immune system may lead to the outgrowth of cancers and increased infectious diseases. EPA should utilize mechanism of action data more fully in their risk assessment.
- Immunotoxicity resulting from woodsmoke inhalation and other sources of human environmental exposure to BaP should be considered by EPA.

- 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; or multiple tissue/cell types; thus, it is important to consider these factors in interpreting temporal- and non-linear dose-relationships for specific immune endpoints.
- The immunotoxicity datasets for benzoapyrene are limited because they utilize rats rather than preferred mouse models, and no sensitive functional assays, such as the T-dependent antibody response (TDAR) were performed. Thymic atrophy is a relatively insensitive endpoint in mice and rats, resulting in a low confidence RfD. Consider whether the composite uncertainty factor addresses the database inadequacies, taking into account the other nonfunctional but relevant immunotoxicity endpoints.
- Developmental immunotoxicity is not well-addressed in the document; it is likely (WHO, 2012) that the developing immune system may be one to two orders of magnitude more sensitive to BaP exposures. Again, consider if uncertainty factors can address the concerns regarding the inadequacy of the database.
- Recommendation: This report could be improved by a well defined, unified approach for immunotoxicity risk assessment (e.g. through a guidance document), that identifies sensitive biomarkers of exposure and effect for the immune system of animals and humans.

Hazard Assessment 2d - Cancer

Scott Burchiel, John DiGiovanni, Helen Goeden, Bhagavatula Moorthy, Miriam Poirier, Kenneth Ramos, Leslie Stayner, Alan Stern

Charge Question on 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?”

Major Conclusions of the Panel on the Carcinogenicity of Benzo[a]pyrene in Humans

- PAH mixtures are carcinogenic to humans in industry (coke oven, aluminum, iron and steel), but because humans are not exposed to benzo[a]pyrene alone, it is not possible to establish causality based on the epidemiology alone. Nonetheless, the epidemiology studies provide strong support for the carcinogenicity of benzo[a]pyrene.
- By the EPA-defined secondary criteria (involving similar mode of action and mechanistic events in humans and animals, tumors in animals, and the likelihood that similar mechanistic events in humans will result in tumor formation) there is sufficient evidence for the carcinogenicity of benzo[a]pyrene in humans.
- A similar overall conclusion has been reached by the International Agency for Research on Cancer (IARC) and Health Canada.
- In humans there is strong evidence that PAH exposures cause lung, bladder and skin tumors, moderate evidence for colon adenomas, and no evidence for liver tumors.

Major Conclusions of the Panel on the MOA of Benzo[a]pyrene

- The Panel agreed that benzo[a]pyrene causes cancer primarily through a mutagenic mechanism. Metabolism via the diol-epoxide pathway links benzo[a]pyrene exposure to carcinogenesis through formation of a stable N2-deoxyguanine adduct, the mutagenic properties of which are well documented. This pathway is considered initiating.
- There are, however, two additional metabolic pathways, the radical cation and the *o*-quinone pathways, which produce multiple types of DNA damage, some stable and some unstable, as well as other non-mutagenic effects.
- Because benzo[a]pyrene is a complete carcinogen, mechanisms beyond mutagenesis, which may arise from the radical cation and *o*-quinone pathways, may also contribute to tumor induction.
- Additional events that may impact the MOA, with an effect on tumor induction, include tissue-specific and interindividual variability in metabolic pathways, as well as cell proliferation, DNA repair, and inflammation.

Suggestions for the content of the EPA Document

The Panel members suggested that the document would be more clear with the following additions:

- A defined documentation of the line of evidence connecting the steps in the diol-epoxide pathway between exposure and tumor formation, with reference to the original classical literature. In particular, a more detailed discussion of the linkage between the diol-epoxide and mutation.
- Acknowledgment that the initiating MOA and promoting MOA of benzo[a]pyrene may involve different mechanisms.
- A listing of factors that might alter the known MOAs, and might therefore result in different tumor outcomes.
- A comparison of the relevant importance for cancer risk of the diol-epoxide, mutagenic MOA, compared to the non-mutagenic and/or indirect mutagenic MOAs.
- A table used to define and clarify the methodologies for DNA adduct assays and what they measure, as well as proper nomenclature when describing different types of DNA damage induced by benzo[a]pyrene.
- A reconsideration of the requirement of human epidemiology studies to have individual monitoring data to be considered Tier 1 studies. Many relevant occupational studies have been omitted from the document for lack of exposure monitoring, but biomarkers such as 1-OH-pyrene and the BPdG adduct, are also good indicators of exposure.
- Additional studies added to Table D-33 that describe benzo[a]pyrene-DNA adducts in humans.
- A table summarizing human epidemiological studies in which human cancer risk has been evaluated in individuals with different levels of PAH-DNA damage, and an Odds Ratio of Relative Risk calculated for those with the highest levels of PAH-DNA damage.

Question 2e. Other types of toxicity (section 1.1.4).

Burchiel, Choi, English, Li, Ramos, Moorthy, Vorhees

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?

- We agree that the available evidence presented does not support liver, kidney, and hematological effects as human hazards, recognizing that EPA's rationale for those conclusions is incompletely described.
- The available evidence presented supports forestomach toxicity. Be internally consistent in the document regarding the human health hazard of forestomach toxicity. Consider factors identified in IARC, 2003 such as mode(s) of action and influencers of target tissue residence time (*viz.*, method and vehicle of benzo[a]pyrene (BaP) administration) in addressing the predictive value for humans of forestomach effects in rodents.

- Further explanation is needed as to the rationale for concluding that the available evidence does not support cardiovascular effects as a potential human hazard.
- The cardiovascular system may be impacted through multiple modes, so integrate human, animal, and mechanistic evidence, e.g., BaP-induced atherosclerosis in mice, induction of inflammatory cytokines and ROS, cardiovascular effects from gestational exposure.
- Further explanation is needed as to the rationale for concluding that the available evidence does not support adult nervous system effects as a potential human hazard. Panel members have provided analyses and interpretations of relevant studies.
- Adult and developmental pulmonary toxicity are not well addressed in the document. Panel members have provided relevant references; e.g., BaP effect on susceptibility of newborn mice to hyperoxic lung injury and chronic lung disease.
- Adult and developmental renal toxicity are not well addressed in the document. Panel members have provided relevant references; e.g. BaP effects on renal function in rats, Intrauterine effects of BaP on kidney morphogenesis and late onset renal disease.

Renal References

- Alejandro, N. F., Parrish, A. R., Bowes III, R.C., Burghardt, R.C. and Ramos, K. S. Phenotypic profiles of cultured glomerular cells following repeated cycles of hydrocarbon injury. *Kidney International* 57(4), 1571-1580, Apr 2000. PMID: 10760092.
- Parrish, A.R., Alejandro, N.F., Bral, C.M., Kerzee, J.K., Bowes, R.C.III and Ramos, K.S. Characterization of glomerular cell phenotypes following repeated cycles of benzo(a)pyrene injury in vitro. *Biochemical Pharmacology*, 64(1), 31-39, Jul 2002. PMID: 12106603
- Nanez, A., Alejandro, N.F., Falahatpisheh, M.H., Roths, J.B. and Ramos, K.S. Disruption of cell-cell and cell-matrix interactions in hydrocarbon nephropathy. *American Journal of Physiology-Renal* 289(6), F1291-F1303, Dec 2005. Epub 2005 Jul 5. PMID: 15998846.
- Valentovic, M.A., Alejandro, N, Brown, P.I. and Ramos, K.S. Streptozotocin (STZ) diabetes enhances benzo(a)pyrene-induced renal injury in Sprague Dawley rats. *Toxicology Letters* 164(3), 214-220, Jul 14 2006. Epub 2006 Feb 7. PMID: 16460892.
- Nanez, A., Ramos, I.N. and Ramos, KS. A mutant allele of AHR protects the embryonic kidney from hydrocarbon-induced deficits in fetal programming. *Environmental Health Perspectives* 119, 1745-1753, 2011. PMID 21803694.

Responses to Charge Question # 3a, Oral Reference Dose

Selection of Developmental Endpoints

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 and specific endpoints upon which it is based warrants additional justification by EPA

- Issues with study design and data analysis in Chen et al. (2012)
 - Negative
 - Potential dam and pup stress from repeated rotation of dams
 - Potential nurturing bias against high dose based on smell and/or behavioral differences especially following gavage doses.
 - Use of LSD post hoc test
 - Although the BMD approach for deriving the POD is not dependent on the specific statistical tests used for group comparisons, the overall weight of evidence and evaluation of this study is based on the original statistical analysis using this test which appears inappropriate.
 - The total number of dams used and timing (e.g. litters redistributed to other dams who gave birth within 24 hrs of each other) to achieve 40 litters of 4 M and 4F divided into 10 litters per track was not described. Presumably, all 40 litters were not born in one day, so the details on how this was achieved, including use of >40 litters initially, so that pups are exactly the same age in each litter are critical information for study design that can impact study outcome and interpretation of data.
 - Positive
 - Adequate numbers of litters (40 litters, 10/dose group) were used
 - Good dose-response
 - Multiple and well characterized tests
 - The subjective tests were conducted with observers blind to treatment level
 - Dose-dependent effects were found on multiple behavioral outcomes

- **Panel recommendations**

- Specifically consider the overall picture of neurodevelopmental impact from all of the neurodevelopmental endpoints from Chen et al. (2012), including plus maze, reflex, locomotor activity and water maze to justify and support the choice of the critical endpoint.
- Reconsider or provide stronger justification for not using escape latency from the water Morris maze, which appears to be the most stable behavioral difference that was repeated 4 days for 2 separate tracks (cohort) of animals. EPA is correct that this effect is not a learning or memory effect due to difference in baseline from day 1, but it is some indication of an effect (even if it's locomotor). EPA should explain how the BMD was calculated for escape latency since there are 4 different days for each track and each sex.
- Consider reproductive outcomes
 - Including Gao et al. (2011) cervical hyperplasia and cervical inflammation
- Better explain the reasons for not modeling immunotoxicity (IgM, IgA) endpoints
- Given significant limitations, Xue et al. (2010) should not appear in Table 2-2

Uncertainty Factor Adjustments

- EPA stated that they applied a full UF animal-human of 10 to Chen et al. (2012) because they did not apply a $bw^{3/4}$ adjustment
 - EPA stated that the allometric $bw^{3/4}$ adjustment is not appropriate for extrapolating from neonate animal to adult humans.
- Panel recommendation
 - As this endpoint is a neurodevelopmental endpoint, the extrapolation in question is from neonatal animal to neonatal human (not to adult human)
 - Therefore, consider application of $bw^{3/4}$ adjustment
 - Per EPA 2011 allometric scaling guidance

- Consider changing the presentation order of the UFs starting with LOAEL-NOAEL... and ending with sensitive human as this is the logical flow when beginning with an animal study
- 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 considered a replacement for the multi-gen), F1 animals, which have been continually dosed, are only assessed for reproductive effects if triggered (Parental generation are only required to be dosed for 2-weeks prior to mating). Therefore, it is questionable that the OECD 443 will provide any additionally useful reproductive information.

- The EPA also justifies the UF-database of 3 based on the lack of a study examining functional neurological endpoints following exposure from gestation through lactation.
 - There were 2 oral studies exposing dams GD 14-17 (McCallister et al. 2008, Sheng et al. 2010), and 2 oral studies exposing dams or pups directly postnatally (Bouayed et al. 2009; Chen et al. 2012 that evaluated functional endpoints. There were additional gestational exposure studies evaluating receptor gene expression.
 - There were no studies with both gestational and lactational studies.
 - The EPA should address the question of whether the absence of this study warrants an additional 3x given the 10x for inter-individual differences that is already included. As part of this deliberation, EPA might consider whether an EPA DNT guideline study and/or extended 1-gen study with a DNT cohort is likely to result in a NOAEL below that of Chen et al. 2012.

Additional Issues

- Given the reproductive, developmental and trans-placental effects of BaP, the panel encourages EPA to ensure that available multi-generational and one-generational effects are, to the extent possible addressed
- When possible, EPA should identify the sensitive sex in a given study and use the sensitive sex for dose-response modeling.

Question 3B. Inhalation Reference Concentration

Schlesinger, McIntyre, Bartell, Foster, Goeden, Stern, Walter

In the current form, the overall RfC is inadequately supported.

- Endpoint (fetal death) is relevant for humans
 - Only one inhalation animal study (Archibong et al 2002) with gestational exposure used for RfC generation
 - Two others identified/should be considered
 - Wu (2003)- gestation exposure from GD11-21 (25, 75, 100ug/m3); birth index impacted [noted in draft document]
 - Archibong (2012)- 2-week exposure prior to mating (50, 75, 100 ug/m3) subsequent effects on cycling/ovulation; number of pups born
- Selected study exhibits weaknesses
 - Noted in group write-up, decreases confidence
- Rationale for not using BMD approach is unclear. Unequal variances and lack of access to original data are not sufficient reason to avoid benchmark dose modeling of the fetal death data. EPA has fit benchmark dose models to epidemiological data summaries with these attributes, and should consider those approaches here.

- Given the particle sizes used in the key study the RDDR adjustment (1.1) does not adequately account for interspecies differences in particle deposition and systemic toxicokinetics. Therefore, EPA's application of an UF of 3 to address residual uncertainty in extrapolating from animals to humans is inadequate.
- EPA's current use of one study (that has weaknesses) for determining the POD, coupled with the uncertainties in UF determination, suggests that derivation of an RfC based on this study may not be possible, even with low confidence.
- Document notes that confidence in the database is low, but confidence in the key study is medium.
 - One could argue that the study is fine as a study, but confidence in its applicability to derive an RfC is low.
 - Rationale to elevate to medium is not justifiable

Recommendations

- EPA should also consider the studies by Wu (Int. J. Devl Neuroscience 21 (2003) 333–346) and Archibong (Reproductive Toxicology 34 (2012) 635–643) which describe effects on birth-index data and mean number of pups born, respectively for POD and RfC calculations.
 - Collectively, these three studies may provide insight on the dose-response; increase confidence in the RfC calculation.
- EPA should explore if these three studies are amenable to BMD approaches
- The generation and application of respective Ufs needs further justification/explanation.

3c. Oral Slope Factor for Cancer

- There was a consensus among the reviewers that the two selected lifetime oral carcinogenesis studies were well done and appropriate for the dose-response modeling used for cancer oral slope factor derivation.
 - Rational for selecting only one study (Beland and Culp, 1998, using female B6C3F1 mice) for slope factor derivation rather than using both studies questioned.
 - Both mouse and rat studies deemed appropriate
 - Mouse study used only female mice
 - If no biological basis exists for choosing the mouse study versus rat study, EPA should consider averaging over both studies (e.g., simple averaging as used in previous oral slope factor derivation, or meta-analytic/Bayesian averaging as recommended in the 2014 NRC Review of IRIS).
 - EPA should better explain and justify the decision to base the current oral slope factor for cancer on a single study, in context of the Guidelines for Carcinogen Risk Assessment (2005) as well as the previous oral slope factor that used averaged data from two different studies.

- Concern about the relevance of the rodent forestomach to human oral carcinogenesis since humans do not have a forestomach.
 - This issue requires more thorough discussion in the oral slope factor section and how concordance between mice and rats for tumors in the forestomach further supports the relevance to humans.

- The multistage Weibull model was appropriate for the dose response modeling and preferable due to incorporation of time-to-tumor data, although there were some comments on whether consideration should be given to fitting other models rather than only using the multistage-Weibull model alone.
 - For example, a comparison of values derived from different methods would be informative and help to further support the use of only the multistage-Weibull model.
 - See Fitzgerald et al, EHP, 112:1341-1346, 2004 (and graph below taken from this article).

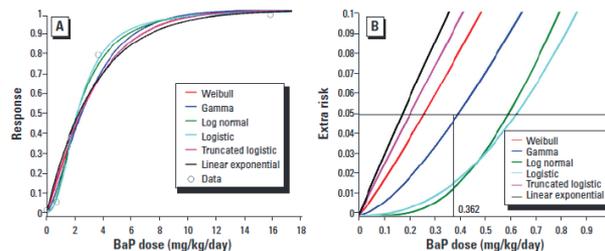


Figure 1. Suite of models fitted to BaP dose-response data (mouse forestomach tumors) reported by Culp et al. (1998). (A) MLE fitting of models except the truncated normal, which could not be fitted. (B) The extra-risk dose curves of (A) in the low-dose region around the 0.05 risk level and averaged dose at 0.362 mg/kg/day.

- The adjustments for approximating human equivalent slope factors used EPAs cross-species allometric scaling methodology of $BW^{3/4}$.
 - Question of whether alimentary tract tumor sites should be scaled using this methodology because they represent portal of entry tissues
 - Comparison of slope factors using older scaling factor of $BW^{2/3}$ compared to the newer scaling factor of $BW^{3/4}$ is suggested to determine the extent to which this revised cross-species scaling adjustment impacted calculation of the newer oral slope factor for cancer.

- A brief description of the derivation of the previous oral slope factor for cancer is given on page 2-32 of the document. It was felt by some reviewers that additional discussion comparing the previous analysis with the current analysis might be useful.

- 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).
 - Suggest that the relevance of this assumption of human exposure should be further discussed.

- Further discussion of how oral exposure to PAH mixtures might influence the carcinogenicity of B(a)P as well as derivation and use of the oral slope factor is suggested.
 - Some discussion of this point should be considered in the Uncertainties section of this report.
 - The study by Culp et al (Carcinogenesis 19:117-124,1998) actually compared the oral carcinogenicity of B(a)P in the two year bioassay with two different coal tar mixtures of known content. The coal tar mixtures produced a lower incidence of forestomach tumors compared to B(a)P but higher incidence of lung tumors. These data were further evaluated and modeled in the publication by Fitzgerald et al (EHP,112:1341-1346, 2004) . Some consideration of these analyses could be used as a starting point for further discussion on this topic.

Question 3d. Inhalation Unit Risk for Cancer (section 2.4)

Choice of Studies

1. Criteria for selection of studies: epidemiologic preferred; animal model respond like human if comparable; route of exposure; lifetime exposure duration; multiple dose levels; and adequate statistical power.
2. Principal animal (male adult hamster) study – Thyssen et al, 1981 : dose exposure (0, 2, 9, and 46 mg/m³ x 4.5 h/d x 10 weeks + 3hr/d x until death).
3. Readouts: body weight, incidence and latency of tumors with segmental distributions (URT, trachea, oro-pharynx, esophagus, forestomach).
4. Main features of the study replicated in subsequent report (Pauluh et al, 1985).

5. Results:

a) comparable to human epi studies wherein lung and bladder cancer associated with PAH occupational exposures (aluminum industries)(Lung CA Risk after expos to PAHs : Review and Meta-analysis, B Armstrong, et al, EHP, 2004).

b) additional confidence in the results of this single study (Thyssen et al, 1981) arises from a subsequent short communication from this same research group (J Pauluhn et al, 1985, Exp Path 28:31, 1985) and although limited in scope results from this 2nd study appear to replicate findings of neoplastic changes in the hamster model exposed to B(a)P aerosol.

6. Issues:

single study, one sex, one species. Panel members suggested references were available for additional animal model studies for EPA to consider in comparison to their reliance upon the single hamster study (Thyssen et al, 1981).

Dose-Response Analysis, Inhalation Unit Risk Derivation, and Uncertainty

- Dose-response methods were appropriate and Multistage Weibull model fit was adequate. Although the panel agrees with EPA that the multistage Weibull model is preferable due to incorporation of time-to-tumor data, supplemental analysis using other dose-response models would be informative and help to further support the use of the unit risk derived from the multistage-Weibull model.
- 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"; p. 2-35, lines 6-8) should be discussed. EPA should address whether these are reasonable assumptions.
- It would be helpful for EPA to address how reasonable it is that lifetime exposures will be in the approximately linear low dose region ($<0.3 \text{ mg/m}^3$, the human equivalent POD).
- EPA should state a conclusion regarding overall uncertainty or level of confidence for the inhalation unit risk (as proposed in the EPA draft handbook for IRIS and endorsed on p. 118 of the NRC 2014 review of the IRIS program).
- EPA should better justify selection of body weight scaling factors in relation to "portal of entry," as discussed in the EPA Guidelines for Carcinogen Risk Assessment.

- Modeling assumptions were reasonable, but supplemental sensitivity analyses would be informative (other assumptions about latency, cross-species scaling of doses, and not eliminating from the analysis all animals without confirmed examination of one or more of the pharynx or respiratory tract tissues).
- Given the extensive human studies of airborne inhalational exposures to PAHs by coke oven, and aluminum smelter workers, the panel recommends that EPA give further consideration to selection of occupational studies (or meta-analysis of occupational studies) to develop unit risk estimate(s) for Table 2-9, in addition to the single hamster study (Thyssen et al, 1981). Although interpretation of the epidemiological evidence is challenging given that exposures were to mixtures of PAHs with poorly understood interactions, a model using relative potency factors and an assumption of dose additivity was reasonably accurate for some PAH mixtures and conservative for others in one investigation (EPA, 1990), and should be considered for adjustment of epidemiological results in estimation of the unit risk attributable to BaP alone.

Question 3e. Dermal slope factor.

The draft assessment proposes a dermal slope factor of 0.006 per $\mu\text{g}/\text{day}$ 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. Analysis of carcinogenicity data (choice of studies) and epi studies, including pharmaceutical coal tar epi studies

2.5.2. Dose-response analysis & dermal absorption and dosimetrics

2.5.3. Derivation of the dermal slope factor

2.5.4. Dermal slope factor cross-species scaling

2.5.5. Uncertainties in the derivation of the dermal slope factor

2.5.1. Analysis of carcinogenicity data (choice of Studies)

Choice of skin cancer bioassay studies for developing the dermal slope factor (DSF)

- BaP document reviewed 10 complete carcinogenicity mouse skin tumor bioassay studies from 1959 to 1997 (summarized in Table 2-11) and Sivak et al., 1997 was chosen as the principal study.
- Other skin cancer bioassay studies are mentioned and excluded for further analysis because: (1) only one BaP level was considered, (2) all dose levels induced 90-100% incidence of tumors, (3) dose applications were 1x/week (Nesnow et al., 1993) or 1x/2 weeks (Levin et al., 1977), (4) dose was delivered in a vehicle that interacted or enhanced BaP carcinogenicity.
- EPA should consider adding the Nesnow et al., 1993 and Levin et al., 1977 studies to Table 2-11.
- EPA should consider combining results from the different studies shown in Table 2-11. This would strengthen the derived DSF.
- Skin cancer bioassay studies that examined only one BaP level or observed 90-100% incidence of tumors are not usable for estimating points of departure (POD). However, consistencies in the observations of these studies with observations from the studies listed in Table 2-11 and used to develop POD and DSF would strengthen the derived DSF.

2.5.1. (Continued)

The EPA review of the epidemiologic evidence for of skin cancer in humans is not sufficiently thorough. They cite evidence of an excess of skin cancer in studies of roofers [Hammond et al. 1976] and workers exposed to creosote treated wood [Karlehagen et al. 1992 and Tornqvist 1986], but these groups work outside and would thus have substantial exposure to UV. They also note that recent studies of chimney sweeps do not demonstrate an increased skin cancer risk [Hogstedt et al. 2013]. They do not cite or discuss some older studies that reported an excess of in skin cancer in destructive distillation of coal, shale oil extraction, and workers exposed to creosote in brick making and wood impregnation [Boffetta et al. 1997]. It would be informative to more thoroughly review the evidence for skin cancer in studies of coke, steel and iron, coal gasification and aluminum workers given their relevance for evaluating the appropriateness of using the mouse based risk assessment model for predicting skin cancer risk in humans.

The review group does not believe that epidemiologic studies of pharmaceutical use of coal tar preparations provide an adequate basis for either hazard identification or the derivation of a dermal slope factor due to uncertainties regarding the PAH dose that results from shampoo use and the relevance of the (psoriasis patient) population.

2.5.2. Dose-response analysis; Dermal absorption and dosimetrics

2.5.3. Derivation of the dermal slope factor

BaP document states that mass rather than mass/area can be used as the appropriate dose metric for cancer risk at “low doses” of BaP. Low dose needs to be defined (see further comments on slide 6).

Choice of dose metric

- Published dermal slope factors for BaP skin carcinogenesis have used mass and mass/skin area as dose metrics
- There does not appear to be any empirical data available to inform a choice between these two dose metrics or to select another
- EPA proposes to use mass as the dose metric, but does not provide a convincing rationale
- The committee does not have a specific recommendation as to dose metric, but strongly recommends that in the absence of empirical data the decision be based upon a clearly articulated, logical, scientific structure that includes what is known about the dermal absorption of BaP under both conditions of the bioassay(s) and anticipated human exposures, as well as the mechanism of skin carcinogenesis of BaP.

2.5.2. & 2.5.3 (Continued)

Experimental studies have demonstrated that equal masses of chemical absorb into the skin when the area of direct chemical contact is less than the applied skin area (i.e., the mass of chemical applied is too small to completely cover the application area).

- E.g., observation from Roy and Singh, 2001 that % of BaP applied on contaminated soil that absorbed was independent of the mass of soil applied until the skin surface area was completely covered with soil; further increases in the mass of soil applied caused % BaP absorption to decrease.

The DSF derived from the skin cancer bioassay in mice is based on the applied dose, which most probably closely approximates the absorbed dose.

- The time between dose applications was long enough and the applied doses small enough in the mouse studies for ~100% absorption (e.g., Wester et al., 1990 observed 51% (*in vivo* monkey) and 24% (*in vitro* human) for 0.5 µg/cm² in 24 h; absorption rates through mouse skin are faster than through humans and monkeys.
- The conclusion that absorbed dose approximately equals the applied dose assumes that dose losses were minimal; study protocols should be evaluated for factors that may have affected losses of the applied dose (e.g. by grooming)

Cancer risk calculated from the derived DSF should use absorbed dose (not exposed applied dose)

2.5.2. & 2.5.3 (Continued)

EPA should describe what constitutes a “low dose” for the assumption that mass of BaP is the appropriate dose metric for calculating the DSF from the skin cancer bioassay studies and for estimating cancer risk in humans

- This should be consistent with the proposed logical structure for skin cancer from skin exposure to BaP, which is a solid at skin temperature.
- For dermal absorption: skin area with direct chemical contact must be less than the total applied area; i.e., mass of BaP applied cannot cover completely the applied area
 - For BaP deposited onto skin from a volatile solvent, the mass of BaP that would give a theoretical uniformly thick film $\leq 1 \mu\text{m}$ (i.e., $\sim 135 \mu\text{g}$ of BaP/cm²) would be too small to completely cover the application area, where:

Theoretical thickness of a uniform film on the application area = [(BaP mass applied)/(application area)]/ ρ_{BaP} ; ρ_{BaP} = density of BaP = 1.35 g/mL

- Metabolism in the target tissue (the viable epidermis) should not be saturated.
- The document identifies the linear limit for using the slope factor to calculate cancer risk in humans based on the human equivalent point-of-departure (POD_{HED} = 17.9 $\mu\text{g}/\text{day}$) estimated from the mouse POD_M adjusted by the mouse-to-human scaling factor as the body weight ratio to the $3/4$ power. This is an appropriate limit that could be smaller than 17.9 $\mu\text{g}/\text{day}$ for different scaling factor approaches.

2.5.2. & 2.5.3 (Continued)

EPA should add diagrams illustrating:

- The steps involved in calculating human cancer risk based on skin cancer bioassay studies in mice; for example
 - Tumors observed in mouse studied as a function of time and exposed dose
 - Exposed dose \approx applied dose to estimate in mice: POD_m and DSF_m
 - DSF_m scaled to the human DSF_h
 - Estimate of absorbed dose from exposed dose and exposure scenario
 - Human cancer risk = $DSF_h \times$ (Absorbed dose)
- The logical structure (physiological steps to carcinogenesis) to facilitate choices of dose metric and cross-species scaling

2.5.4. Dermal slope factor cross-species scaling

Starting point is dermal scaling factor in the mouse (i.e., $DSF_m = 1.7$ (ug/day)¹), which is adjusted by the appropriate human to mouse ratio to obtain the dermal slope factor in humans (DSF_h)

Experimental cancer risk information for scaling from mouse to human skin cancer from dermal exposure is not available.

It is unknown if the chosen approach for scaling of skin cancer risk from BaP exposure to skin is similar to interspecies differences in whole body toxicokinetics, which is the approach adopted by EPA.

Alternative approaches for scaling are listed. The science for choosing the best approach is uncertain. Therefore, the chosen scaling approach should be supported by a coherent logical structure.

Differences between mouse and human skin should be considered in light of the proposed logical structure for skin cancer risk; for example:

- Thickness of and metabolic rates in the target tissue (i.e., the viable epidermis layer)
- Differences in stratum corneum thickness will affect the absorbed dose from a given exposed dose applied to humans compared with mice. However, it may not affect the cross-species scaling of the DSF, which is based on absorbed dose.

2.5.5. Uncertainties in the derivation of the dermal slope factor

The cross-species mouse-to-human scaling of the DSF is a significant contributor.

Other recommendations for describing cancer risk calculated with the DSF

The cancer risk calculation in mice (and therefore in humans) depends on absorbed dose; i.e., $\text{Cancer Risk} = \text{DSF} \times (\text{Absorbed dose})$

EPA should state clearly how the absorbed dose estimates from exposed dose enters the calculation of cancer risk

In actual BaP exposures (from soil or other environmental media), the absorbed dose should be estimated from the exposed dose and the exposure scenario.

- A soil-to-acetone absorption ratio as described in the response to public comments is unnecessary.
- Cancer risk from BaP on soil should be calculated from the estimated absorbed dose from exposure to BaP contaminated soil

Examples of cancer risk estimates from exposure to BaP contaminated soil will use an estimate of the absorbed dose taken from the literature (or RAGS, Vol. 1, Part E). Because the document does not critically review this literature,

- The literature of dermal absorption measurements from BaP contaminated soils be listed
- This estimate of absorption used in the risk calculation should be identified as an example (and not an endorsement of the value used)

Each environmental media will have its own absorption characteristics that should be considered in estimating an absorbed dose for estimating cancer risk

CQ3f. Age-Dependent Adjustment Factors

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. Do the available mechanistic studies in humans and animals support a mutagenic mode of action for cancer induced by benzo(a)pyrene?

The available mechanistic studies in humans and animals support a mutagenic mode of action for BaP-induced cancers. Given that the EPA/630/R-03/003F “Supplemental Guidance for Assessing Susceptibility from Early-Life Exposures to Carcinogens” establishes a rational approach for the adjustment of tumor risk for exposures at different ages for carcinogens with a mutagenic mode of action, the committee concludes that the proposed use of age-dependent adjustment factors (ADAFs) is justified.

Charge Question 4 – Executive Summary

- *CQ4: Does the Executive Summary clearly and appropriately present the major conclusions of the assessment?*
- Major conclusions were clearly and adequately presented in the ES
- Suggestions:
 - Shaded box is intended to be a lay language abstract. It should be clearly identified as such, stand alone outside the ES (because it has a different audience), and be examined to insure that the language level is appropriate for its audience. The Panel noted that some language in the current version may not be understandable to lay audiences.
 - Add introductory text to ES that explains why BaP assessment is important in evaluating hazard and risk to human PAH exposures, that are always to mixtures
 - While it is important to capture important conclusions of the assessment, the Agency should strive to accomplish this in a readable length.
 - Adding a few sentences on how confidence (e.g., “medium”) is determined would be helpful

Charge Question 5 – Appendix G

- *CQ5: 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.*
- Most scientific issues raised by the public, as summarized in Appendix G, were adequately addressed.
- Panel urges greater transparency in how public comments are distilled into a list of scientific issues meriting an EPA response in the Toxicological Review.
Suggest:
 - Providing a short description of the process of deciding which comments to include in a public response appendix and how comments are aggregated
 - In each Toxicological Review, provide a table that shows the distilled topics and which commenters provided comments on each.

Comments on specific EPA responses

- Topic: Metric to characterize results in the elevated maze (pg G-5)
 - Panel opinions provided in response to CQ2a.
- Topic: Anxiety-like effects as a critical effect (pg G-6)
 - Panel opinions provided in response to CQ2a.
- Topic: Cross species extrapolation of dermal slope factors (pg G-11)
 - Panel opinions provided in response to CQ3e.
- Topic: Appropriate dose metric for BaP dermal carcinogenicity (pg. G-12)
 - Panel opinions provided in response to CQ3e.
- Topic: Appropriate dermal bioavailability of BaP from soil (pg G-12)
 - Panel opinions provided in response to CQ3e.
- Topic: Ground truthing calculations for dermal cancer slope factor (G-12)
 - Panel supports ground truthing exercise for proposed tox values; consistent with NRC recommendations.
 - Limitations were noted in estimated risk versus observed cancer incidence comparisons by both public commenters and EPA in their response (e.g., did not address substantial under-reporting of skin cancer).
 - Panel suggests an improved ground truthing exercise where comparisons are made in the context in which the cancer slope factor will be used.