

**Comments on the U.S. EPA IRIS Toxicological Review of
Benzo[a]pyrene (Public Comment External Review Draft).
U.S. Environmental Protection Agency, Washington, DC,
EPA/635/R-13/138a-b, 2013.**

**American Coke and Coal Chemicals Institute
American Fuels and Petrochemical Manufacturers
American Petroleum Institute
Asphalt Institute
Association of American Railroads
Beazer East, Inc.
Pavement Coatings Technology Council**

November 2013

Table of Contents

I. Summary Comments	1
II. Specific Comments	8
General Charge Questions	8
Chemical-Specific Charge Questions	9
A. Executive Summary	9
B. Literature Search Strategy/Study Selection	10
C. Hazard Identification	10
D. Dose-Response Analysis	43
A. Lack of Real World Validation	113
B. Lack of Discussion of Policy Implications	118
III. References	120

Acronyms and Abbreviations

AIC	Akaike's Information Criterion
BaP	benzo[a]pyrene
BaP-TE	benzo[a]pyrene toxic equivalents
BMCL	benchmark concentration low
BMD	benchmark dose
BMDL	Benchmark Dose Low
BMDS	Benchmark Dose Software
cm ²	square centimeter
DMBA	9,10-dimethyl-1,2-benzanthracene
DMSO	dimethyl sulfoxide
DSF	dermal slope factor
EPM	elevated plus-maze
IARC	International Agency for Research on Cancer
IRIS	Integrated Risk Information System
IUR	inhalation unit risk
LOAEL	Lowest Observed Adverse Effect Level
MCA	methyl cholanthrene
mg	milligram
mg/kg	milligrams per kilogram
mL	milliliter
MTD	maximally tolerated dose
NCEA	National Center for Environmental Assessment
ng	nanogram
ng/g	nanograms/gram
nmol	nanomol
NMSC	non melanoma skin cancer
NOAEL	No Observed Adverse Effect Level

Acronyms and Abbreviations

NRC	National Research Council
NTP	National Toxicology Program
OECD	Organisation for Economic Cooperation and Development
OSF	oral slope factor
PAH	polycyclic aromatic hydrocarbon
PND	postnatal day
POD	points of departure
ppm	parts per million
PUVA	8-methoxypsoralen photochemotherapy
RfC	reference concentration
RfD	reference dose
RPF	relative potency factor
RSL	Regional Screening Level
SD	standard deviation
SEM	standard error of the mean
SMR	standardized mortality ratio
TPA	12-O-tetradecanolyphorbol-13-acetate
UF	uncertainty factor
USEPA	United States Environmental Protection Agency
µg	microgram
µL	microliter
UV	ultraviolet
UVB	ultraviolet-B

I. Summary Comments

General Comments: EPA's *Toxicological Review of Benzo[a]pyrene (CASRN 50 32 8) In Support of Summary Information on the Integrated Risk Information System (IRIS)* (EPA, 2013)does not properly identify the hazards associated with benzo(a)pyrene (BaP), does not properly summarize the literature on PAH-containing mixtures, and incorrectly associates BaP as the sole or principal causative agent in complex mixture toxicity studies.

EPA (2013) is neither clear, concise, nor easy to follow as recommended by NRC (2011), because much of the information in the main document does not match the information provided in the Supplemental Information document. The document structure is not ideal if different information is presented in different sections. Under such conditions, it is not at all clear what the "key outcomes of each step" are.

"Transparent discussions of weight of evidence" as recommended by NRC (2011) are not presented because entire key areas of literature are essentially omitted. In addition, EPA (2013) ignores the literature on ultraviolet light as the major source of human skin cancer by erroneously concluding, instead, that BaP is a major cause of human skin cancer.

EPA's literature search was deficient, because it omitted the entire literature on coal tar pharmaceutical use, which clearly demonstrates that BaP-containing coal tar pharmaceuticals do not cause skin cancer in humans. The literature was provided to EPA, who added several sentences that casually referred to just three of the many meaningful studies provided, thereby giving no serious consideration to its own weight of evidence evaluations. EPA's literature search also omitted the entire literature on human-mouse skin xenografts, which demonstrates that BaP and other PAHs do not cause skin cancer in functioning, viable human skin when skin cancer is seen in the adjacent mouse skin.

Hazard Identification: The Hazard Identification in EPA (2013) has not "clearly and appropriately synthesized" the data for each toxicological effect, because of the over reliance on information about complex mixtures and historical data from years ago when industrial hygiene was not practiced in industry. Humans are not exposed to BaP alone. By its own admission, EPA agrees that BaP is not used in commerce or emitted into the environment from industrial processes. Instead, industrial processes and natural processes produce and use complex mixtures which contain PAHs and hundreds of related and unrelated compounds. Each of these mixtures has a unique composition with greater or lesser quantities of specific PAH compounds, like BaP.

It is, thus, scientifically inappropriate to base the human health risk assessment of hundreds of differing complex mixtures (oils, fuels, coal tars, coal tar pitches, petroleum and coal cokes, petroleum pitches, gasoline and diesel exhausts, boiler emissions, fires, etc.) on the basis of one PAH, BaP, which happens to be well-studied because it was the first PAH that was available for study in the early twentieth century.

The document's hazard identification summary of human skin cancer provides only negative or inconclusive information. Of the twelve studies reported by EPA (2013) as evidence that BaP causes skin cancer in humans, 7 are negative studies and 5 are irrelevant or inconclusive studies. One was a historical review that does not report any epidemiological evidence. The historical summary paper correctly reports the well-known case study literature on skin cancer in chimney sweeps in England prior to the mid-20th century, but these case reports were unique to England and are not relevant to identifying hazards in the 21st century.

EPA (2013) further cited IARC reports on selected complex mixtures or occupations as the basis for a judgment that BaP has been shown to cause skin cancer in humans, but only two of the citations in the cited IARC documents are epidemiology studies. Both of these reported no increase in skin cancer risk. There are at least two additional negative meaningful epidemiological studies that were not cited. The remaining 8 papers cited by IARC are historical case studies that are irrelevant to identifying cancer hazards in the 21st century.

The document's hazard identification summary of human lung cancer provides little useful information. Of the 3 Tier 1 studies reported by EPA (2013) as evidence that BaP causes lung cancer in humans, one is negative, one is inconclusive because of methodological issues with the paper, and one was positive. Other mixture studies were cited, but there is no evidence reported in those studies that links BaP exposure to human lung cancer. Of 9 studies cited, 4 were not significant, 4 were significant, and 1 did not perform statistical significance calculations. In some, but not all studies, workers in certain high temperature environments, such as aluminum production and coke oven facilities, have been shown to have increased rates of lung cancer compared to the general population, but the role of BaP versus the other hundreds of chemicals to which these workers were exposed is unexplained.

EPA (2013) further cited IARC reports on selected complex mixtures or occupations as the basis for a judgment that BaP has been shown to cause lung cancer in humans. Despite the fact that BaP is not listed as a causal agent in these studies, the studies themselves provide only marginal evidence that lung cancer was increased in these workers who were exposed to complex mixtures of chemicals. In fact, there are twice as many studies cited showing no increased risk of lung cancer than there are positive studies. Specifically, of 38 cited studies that were reported to be adequate in quality, 25 were negative studies and 13 were positive studies.

Oral Reference Dose (RfD): It is recommended that EPA revisit the Oral Reference Dose derivation and abandon its reliance on the study of Chen et al. (2012). EPA (2013) reviewed a large number of available studies, rejected many, derived candidate Point of Departure (POD) doses for several, and then rejected all but the POD from Chen et al. (2012) for one endpoint out of over 50 that were measured. The Chen et al. (2012) study is problematic because it involved so many comparisons between control and test animals that statistical significance by chance cannot be ruled out.

In addition to the statistical concerns, Chen et al. (2013) cannot be considered a properly designed and executed study for several key reasons:

- The Elevated Plus Maze test is documented to be a subjective test the results of which are highly influenced by housing of animals, the scoring method used, the construction of the maze, and pre-test manipulation (Hogg, 1996).
- Housing of animals and periods of resting varied among the animals. Differing periods of resting would have affected the rats' anxiety states, which are the subject of the test.
- The number of open arm entries is a poor metric for increased or decreased anxiety state. According to Hogg (1996), "expression of the open arm data as percentages of the total number of arm entries (to give % number of open arm entries; %no) or total time spent (to give % time on open arms; %t) on either the open or closed arms corrects for overall changes in exploration of the maze and helps to reduce activity-induced artifacts." Chen et al. (2012) did not normalize the data as discussed above and instead reported the raw data as their metric.
- No information was provided on maze construction.
- No information was provided on pre-test manipulation.
- There was high variability in the results of the control animals.

The Elevated Plus Maze is a test used to measure anxiety in rats. Rats that are anxious avoid entry into the open arms of the maze, so increased entries into the open arm measures a reduction in anxiety. EPA (2013) reports the same, that the results in the 70 day old females showed "decreased anxiety-like behavior." Given that this test is used in pharmacology laboratories to test the efficacy of anxiety reducing drugs, it is unclear why decrease in anxiety is labeled an "adverse effect." An RfD for BaP should not be based on the definition of this effect in one sex at one time point as an "adverse effect."

EPA should reject the Chen et al. (2012) study and rely on the total scientific weight of evidence of the other studies presented in the document as noted in the detailed comments. Because BaP is arguably the most, well-studied chemical in the history of toxicology, there is no need for a Database Uncertainty Factor when deriving the RfD. Using the scientific weight of evidence of one developmental study, four reproductive studies and two immunological studies, a reasonable RfD would be 1.6×10^{-3} mg/kg-day.

Reference Concentration (RfC): It is recommended that EPA revisit the Reference Concentration (RfC) derivation. The RfC should not focus on one adverse effect in the Archibong et al. (2002) study, but instead consider the scientific weight of evidence from that study and the companion study from the same laboratory, Wu et al. (2003). It is important to quantitatively consider both studies from the same laboratory because the results from the two studies are complementary and reduce uncertainty in dose-response assessment, if used together. In Archibong et al. (2002), the LOAEL for pup survival was 25 ug/m^3 . The NOAEL was undefined and estimated as 2.5 ug/m^3 . However, in the same laboratory one year later, the NOAEL for pup survival was reported as 25 ug/m^3 . Clearly, the NOAEL for pup survival is somewhere in between 2.5 ug/m^3 and 25 ug/m^3 and both studies should be used in the quantitative

dose-response assessment. As noted in the detailed comments, a more reasonable RfC that quantitatively considers the scientific weight of evidence is 1×10^{-4} mg/m³.

Oral Slope Factor (OSF): The commenters agree that the studies of Kroese et al. (2001) and Beland and Culp (1998) are the best available studies for the assessment of the carcinogenic potency of BaP in rodents. The commenters disagree, however, that forestomach tumors are relevant to the assessment of human cancer risk, because humans lack a forestomach. When this criticism has been made in the past, EPA has responded that its cancer assessment guidelines do not require tumor site concordance. EPA has also commented that esophageal tissue is similar in nature to rodent forestomach tissue. In this particular case, esophageal tumor results were observed, and dose-response modeling can be performed directly on esophageal tumor data. Given that EPA normally uses rodent forestomach tumor risk as a surrogate for esophageal tumor risk in humans, the commenters recommend that this approximation step be omitted. Instead, EPA (2013) should directly model the esophageal tumor incidence in rodents and use those results to make estimates of human risk in esophageal tissues. The resulting OSF using the linear extrapolation method is $0.2 \text{ (mg/kg-day)}^{-1}$.

However, the low dose linear extrapolation is not scientifically supported. Plots resulting from EPA's Benchmark Dose Modeling Software show that the data clearly exhibit a threshold and not a linear response near the origin. In fact, esophageal, tongue, and larynx tumor incidences are 0% in the 5 ppm dose group, explicitly demonstrating a threshold for these effects. Therefore, non-linear dose response modeling should be used to derive an OSF for BaP.

Inhalation Unit Risk (IUR): The Unit Risk (IUR) proposed in EPA (2013) must be abandoned because of a number of problems and issues in study design, study reporting, and EPA's dose-response modeling. First, EPA (2013) ignored a companion study from the same laboratory (Pauluhn et al., 1985) that contradicts Thyssen et al. (1981). Pauluhn et al. (1985) is only a short abstract, but EPA should have obtained the raw data from the authors in the same way that they obtained the raw data from the 1981 study, which is a short communication which also shows insufficient data for dose-response modeling. It is unacceptable to rely on one positive study and totally disregard a negative study especially when the negative study uses the same investigators, the same laboratory, the same animals, and the same BaP aerosol generation methods.

A major shortcoming of Thyssen et al. (1981) is that the highest dose exceeded the Maximally Tolerated Dose with high mortality due to particle overload and is thus inappropriate for dose-response modeling. The use of a salt aerosol of unknown concentration also renders the study unusable for dose-response modeling.

There is also considerable confusion about the dosage, the number of animals and the number of tumors in the Thyssen et al. (1981) study. Comparison of the Thyssen paper, the EPA (1990) report, and different sections of the EPA (2013) document shows differing numbers of animals and differing numbers of tumors depending on the report section and table. These are two key inputs into dose-response modeling, and an IUR on the IRIS database cannot be based on an uncertain dataset.

Another key element of dose-response modeling is the *dose*. According to EPA (1990): "Data exist for exposure [sic] measured in the actual exposure chambers over the three years that the entire experiment was conducted. Variability of the measurements over time from the corresponding nominal value was apparent. Also, the duration and frequency of exposure varied among time segments of the experimental period, and animals were exposed only in segments of the entire experimental period." Aside from this unusual situation, there was great variability in the lifetime average exposure received by each animal. EPA (2013) acknowledged this fact, but used the average lifetime exposure for all animals in an exposure group despite the variability of 26%, which exceeds OECD guidelines.

Careful analysis of the data from Thyssen et al. (1981) reveals that 80-90% of the animals that developed tumors received higher than the nominal average dose that was modeled by EPA (2013), so the calculated IUR is biased high.

Many other issues render this study unsuitable for dose-response assessment, so EPA should abandon its use. Moreover, this study should not be used in a manner that assumes low dose linearity. Regretably, EPA (2013) used the standard default method of assuming linearity even though tumor incidence of the lowest dose group was 0%, demonstrating clear evidence of a threshold below which no cancer effects are seen.

If the IUR is not abandoned entirely, it is recommended that EPA clear up the uncertainties associated with the numbers of animals and the numbers of tumors. A scientifically credible IUR cannot be based on a study for which these facts are unclear. The dose-response data should be re-modeled with group doses that match the doses received by the animals that developed tumors rather than the average doses for the groups, which are considerably lower than the doses the affected animals received. Lastly, dose-response modeling should not use linear extrapolation methods that ignore that obvious threshold in the laboratory results.

Dermal Slope Factor (DSF): The DSF proposed in EPA (2013) is the most problematic and must be abandoned, because of a large number of issues that are described in the detailed comments below. At a policy level, the DSF should be abandoned, because there is little evidence that humans are at increased risk of developing skin cancer following dermal exposure to BaP. EPA (2013) overlooked the extensive literature on pharmaceutical users of FDA-approved coal tar ointments and salves. These populations have been studied repeatedly over the years by epidemiologists to determine if they contract skin cancer at higher rates than the general population, and the studies overwhelmingly and consistently show that they do not. EPA (2013) also overlooked the skin cancer experimental literature that directly compares human skin to mouse skin. These studies demonstrate that mouse skin is sensitive to BaP-induced skin cancer and human skin is not. The persuasive studies have grafted human skin onto mouse backs and then dosed both with BaP and other PAHs known to cause skin cancer in the mouse. These studies have repeatedly and reproducibly shown that the functioning human skin does not develop skin cancer as does the mouse skin beyond the margins of the grafts.

Based on the fact that human skin does not behave like mouse skin when treated with BaP, it is unnecessary and incorrect to derive a DSF at all. In fact, FDA recently reviewed the literature on skin cancer and concluded that BaP-containing coal tar products applied dermally to human skin did not increase the human risk of skin cancer. If pure coal tar is safe and efficacious as a dermally applied pharmaceutical product, it is illogical for EPA to regulate incidental and low level dermal contact to coal tar and its component PAHs with a high potency DSF.

In addition to the global comment noted above, there are many technical flaws and errors in the DSF calculation that must be addressed if EPA does not abandon the DSF entirely. These include:

- Dermal dosimetry in skin painting studies is not amenable to traditional dose-response assessment;
- Relevant studies were omitted;
- Key studies had inadequate and poorly defined dosimetry;
- Key studies exceeded the maximally tolerated dose and failed to meet EPA criteria for dermal studies;
- Doses were inappropriately averaged over periods when mice were dead;
- All but two key studies were dismissed and the DSF was derived from the two studies of poorest quality; and
- EPA's Benchmark Dose Modeling criterion for goodness of fit of data was ignored and a less stringent criterion was used.

The DSF uses average daily dose as a metric for dose-response modeling as is typical for an ingestion or inhalation study. However, dermal exposures are known to be subject to the "depot" effect which causes the doses received over time to increase. Thus, the relevant dose of BaP in the skin tissue was not the dose given on any specific day. The true and relevant dose to the skin is cumulative and increases over time. Erroneously using daily dose to derive a DSF results in a meaningless DSF that is artificially high.

Relevant studies were not considered, including a study from EPA's own laboratories. The omitted studies yield candidate DSFs that are lower than those presented in EPA (2013).

None of the studies summarized in EPA (2013) pass even the minimum criteria for data quality (i.e. Klimish score). For instance, the two studies that EPA chooses for the DSF have the following shortcomings:

- BaP source and purity not defined;

- BaP concentration not verified;
- Delivered dose not quantified; and
- Skin surface area receiving dose not defined.

EPA's own acceptance criteria for dermal studies were not followed for one of the two studies on which the DSF is based. EPA (1988) has defined the Maximum Tolerated Dose for a dermal study as a dose that does not cause a "marked inflammatory response or ulcerative lesion." Sivek et al. (1997) clearly reported skin lesions inconsistent with these criteria.

EPA (2013) also performed an unsupported data adjustment by averaging the daily doses given over the period of the experiment over an arbitrary 104 weeks even when animals in many studies were dead long before this time. For instance, Poel (1959) used C57L mice, which normally live less than 70 weeks. EPA (2013) calculated the daily dose that these C57L mice would have received as if they had been some other strain of mouse and lived to 104 weeks. However, these mice did not live to 104 weeks. They were all dead after ~60 weeks. There is simply no logic to this "data adjustment." The implications, however, are important. By artificially reducing the dose that was entered into the benchmark dose modeling software, the DSF was biased upward. There is no scientific rationale for this "data adjustment," and it must be reversed.

EPA (2013) also deviated from its own guidance regarding the statistical significance level used to determine which model fits were "acceptable fits" to the data. This *alpha* value is listed in EPA guidance, is standard practice, and was used elsewhere in the 2013 document. However, for the DSF, EPA (2013) inexplicably deviated from its stated guidance and used a more lenient *alpha* value, thus classifying model results as appropriate fits, when they were not. EPA must re-model these data sets using the appropriate statistical criterion.

In summary, EPA (2013) assessed the literature, chose certain studies while rejecting others of equal or superior quality, chose selected data from individual papers rejecting other data from the same paper for unknown or for unexplained reasons, derived candidate Point of Departure doses and candidate DSFs from selected datasets, and then inexplicably rejected all but two of these doses. This manner of selectively using the available data has not been uncommon in the past in the derivation of toxicological criteria values. However, this approach of focusing always on worst case datasets does not result in a dose-response criterion that reflects the full scientific weight of evidence on a particular chemical and hazard. Taking the full scientific weight of evidence into account is quite simple. It is recommended that EPA identify all studies that meet data quality criteria and then take the average of the candidate DSFs as the proposed DSF. Such an approach is scientifically neutral, takes the full scientific weight of evidence into account, and reduces uncertainty by using data from different studies done in different laboratories with different strains of mice.

Lack of Real World Validation: EPA (2013) has not performed any validations to determine if the proposed DSF is logical and could possibly be true. EPA (2013) has derived a DSF for BaP that, if

finalized in its present form, would drastically reshape the manner in which presumed hazardous site investigations and subsequent remediation are performed. If this DSF was a true predictive indicator of human skin cancer risk, it would mean that ingesting BaP and BaP-toxic equivalents would pose much less risk to human health than incidental dermal contact. It would mean that 100% of users of pharmaceutical coal tar products *should* have skin cancer, when, in fact they do not. It would also mean that BaP and BaP-toxic equivalents in soil throughout the US are the cause of 30% of all human skin cancer, which cannot possibly be true. For these reasons, it is therefore recommended that EPA abandon the DSF entirely, but if it does not, then any future proposed DSF should be subjected to a *real world validation* to determine if the DSF is scientifically supportable.

II. Specific Comments

General Charge Questions

1. NRC (2011) indicated that the introductory section of IRIS assessments needed to be expanded to describe more fully the methods of the assessment. NRC stated that they were “not recommending the addition of long descriptions of USEPA guidelines to the introduction, but rather clear, concise statements of criteria used to exclude, include, and advance studies for derivation of [toxicity values].” Please comment on whether the new Preamble provides a clear and concise description of the guidance and methods that USEPA uses in developing IRIS assessments.

COMMENT: No comment due to time constraints. No comment does not mean tacit agreement.

2. NRC (2011) provided comments on ways to improve the presentation of steps used to generate IRIS assessments and indicated key outcomes at each step, including systematic review of evidence, hazard identification, and dose-response assessment. Please comment on the new IRIS document structure and whether it will increase the ability for the assessments to be more clear, concise, and easy to follow.

COMMENT:

It is noted that much of the information in the main document does not match the information provided in the Supplemental Information document. Specifically, many of the dermal Point of Departure doses (Benchmark Dose Low₁₀) do not match. The document structure is not ideal if different information is presented in different sections. Under such conditions, it is not at all clear what the “key outcomes of each step” are.

3. NRC (2011) states that “all critical studies need to be thoroughly evaluated with standardized approaches that are clearly formulated” and that “strengthened, more integrative, and more transparent discussions of weight of evidence are needed.” NRC also indicated that the changes suggested would involve a multiyear process. Please comment on USEPA’s success thus far in implementing these recommendations.

COMMENT:

“Transparent discussions of weight of evidence” are not presented because entire key areas of literature are not discussed at all. These include the coal tar pharmaceutical epidemiology and the human-skin xenograft literature. In addition, USEPA (2013) ignores the literature on ultraviolet light as the major source of human skin cancer when concluding, instead, that BaP is the major cause of human skin cancer. In addition, USEPA (2013) focused throughout the document on one or two studies, excluding all others, instead of utilizing the full scientific weight of evidence. In its review of the draft IRIS assessment of formaldehyde, the National Research Council (NRC, 2011) has endorsed and instructed USEPA to take into account the full scientific weight of evidence when deriving toxicological criteria, rather than focusing on a single study:

“In principle, identifying the ‘best’ study for general risk assessment purposes is neither feasible nor necessary. Inclusion of multiple studies that meet the selection criteria will enhance EPA’s ability to examine variability and uncertainty attributable to, for example, different study designs, populations, and exposure conditions.”

“Overall, the committee found little synthesis of the relationships among the identified noncancer health effects; it appeared that EPA was driven by the need to identify the best study for each health effect rather than trying to integrate all the information.”

“The NRC Committee to Review EPA’s Toxicological Assessment of Tetrachloroethylene (NRC, 2010) made several recommendations for advancing methodology and promoting applications. Further research is needed to study various approaches. Small (2008) discussed a probabilistic framework. Given a set of options related to a key assumption (such as mode of action) or a key choice (such as cancer end point), a preference score (or prior probability) may be assigned to each option. The final risk estimate thus also has a weight or probability attached that combines the preference on all options over each assumption or choice. The overarching weight is the result of propagation of uncertainty in each assumption or choice and aggregation of all assumptions over the risk assessment process tree. The collection of final risk estimates for all permissible combinations of assumption and choice forms an empirical distribution. That distribution quantifies the full range of variation and uncertainty in the risk estimate. With the full range of variation of risk estimates and other information on preference of key assumptions and choices, regulatory policy can depend less on a single principal study, a single principal dataset, or a principal end point. The risk-management process may use the distributional properties of the risk estimate to choose a final risk estimate in the context of all feasible assumptions and choices.”

Chemical-Specific Charge Questions

A. Executive Summary

1. The major conclusions of the assessment pertaining to the hazard identification and dose-response analysis have been summarized in the Executive Summary. Please comment on whether the

conclusions have been clearly and sufficiently described for purposes of condensing the Toxicological Review information into a concise summary.

COMMENT: No comment due to time constraints. No comment does not mean tacit agreement.

B. Literature Search Strategy/Study Selection

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

COMMENT:

USEPA's literature search omitted the entire literature on coal tar pharmaceutical use, which demonstrates that BaP-containing coal tar pharmaceuticals do not cause skin cancer in humans. The literature was provided to USEPA, and several sentences were added that casually referred to three of the many studies provided, but the information was not seriously considered.

USEPA's literature search omitted the entire literature on human-mouse skin xenografts that demonstrates that BaP and other PAHs do not cause skin cancer in functioning, viable human skin, but skin cancer is seen in the adjacent mouse skin.

C. Hazard Identification

Synthesis of Evidence

1. A synthesis of the evidence for benzo[a]pyrene toxicity is provided in Chapter 1, Hazard Identification. Please comment on whether the available data have been clearly and appropriately synthesized for each toxicological effect. Please comment on whether the weight of evidence for hazard identification has been clearly described and scientifically justified.

COMMENT:

Inappropriate Focus on Benzo(a)pyrene

The entire foundation of this IRIS assessment is that BaP is the only, the major, or the most potent PAH of interest in all of the hundreds of naturally occurring and man-made mixtures of PAHs. It is not. It *is* the most studied PAH, but this is mere happenstance. As described in *Fifty years of benzo(a)pyrene* (Phillips, 1983), BaP has been the focus of so much research because it was the first PAH to be isolated and crystallized from coal tar in the early part of the twentieth century. In 1930, J.W. Cook distilled and fractionated two tons of coal tar pitch, and in 1931 he and his co-workers isolated “about 7 g of a yellow crystalline material of melting point 116°C.” Later in 1931, Cook isolated two pure substances from this material, the major component melting at 176°C and the minor component at 187°C. Two chemicals were synthesized and found to be identical to the crystals isolated from coal tar. They were BaP and benzo(e)pyrene (BeP). Samples of pure BaP were now available for study, and they were widely distributed. BaP now has a vast literature because it was the most easily obtainable PAH to study, and because of the mounting literature on BaP to which researchers could compare their results.

Humans are not exposed to BaP alone. BaP is not used in commerce or emitted into the environment from industrial processes. Instead, industrial processes and natural processes produce and use complex mixtures which contain PAHs and hundreds of related and unrelated compounds. Each of these mixtures has a unique composition with greater or lesser quantities of specific PAH compounds, like BaP. Even a seemingly well-understood mixture like coal tar is not unique. Coal tars differ in composition depending on the source of the coal, the configuration of the coke oven or manufactured gas plant in which they are formed, and the time, temperature, and pressure conditions to which it is exposed.

It is, thus, scientifically inappropriate to base the human health risk assessment of hundreds of differing complex mixtures (oils, fuels, coal tars, coal tar pitches, petroleum and coal cokes, petroleum pitches, gasoline and diesel exhausts, boiler emissions, fires, etc.) on the basis of one PAH, BaP, which happens to be well-studied because it was readily available for study.

In addition to the fact that the different mixtures have different compositions of PAHs which on their own in isolation might have a certain risk profile, each complex mixture has a different composition of PAH compounds that antagonize each other (API et al., 2010).

Skin Cancer in Humans

USEPA (2013) has mischaracterized the weight of evidence that BaP causes skin cancer in humans.

USEPA (2013) provides a discussion of the evidence that BaP causes skin cancer in humans by reporting several studies of workers in industries who were likely exposed not only to BaP, but also other PAHs, and many other chemical and non-chemical agents, such as ultraviolet light. Of the twelve studies reported by USEPA (2013) as evidence that BaP causes skin cancer in humans, 7 are negative studies and 5 are irrelevant or inconclusive studies. One was a historical review that does not report any epidemiological evidence. The historical summary paper correctly reports the well-known case study literature on skin cancer in chimney sweeps in England prior to the mid-20th century, but these case reports are not relevant to identifying hazards in the 21st century.

On page 1-52, the document presents the evidence that BaP may cause melanoma in humans but ultimately concludes that there is no evidence that it does. Spinelli et al. (2006) and Gibbs et al. (2007a, 2007b) are correctly reported as showing no statistically significant increase in melanoma in aluminum reduction workers.

Next, USEPA (2013) cites Brown and Thornton (1957) but provides no data. Brown and Thornton (1957) is a historical summary paper that discusses the scrotal cancers in British chimney sweeps reported in a 1755 medical lecture by Percival Pott. The article also reports several case studies on chimney sweeps: Henry (1946), Hueper (1942), Kennaway (1925), and Kennaway and Kennaway (1937). These are not epidemiological studies, and are case reports. Brown and Thornton (1957) discuss the generally recognized fact that British chimney sweeps prior to the mid-20th century did have high rates of scrotal cancer, but that this was seen in England and not elsewhere. Scrotal cancers were not seen in chimney sweeps in other European countries or the United States. The reasons stated by Butlin (1892) are several: different type of coal, lack of protective clothing and lack of regular bathing. Whatever the reasons, scrotal cancer in chimney sweeps has not been observed since before the middle of the 20th century.

USEPA (2013) then mis-cites the work of Hammond et al. (1976) by reporting a statistically significant increase in non-melanoma when no statistically significant increase was seen. USEPA (2013) reports that this citation was a study of asphalt workers (roofers). In fact, the authors state that the workers were exposed to pitch and asphalt. The study population was not just roofers, as USEPA reports. It was a study of roofers and water proofers. USEPA (2013) reports that workers with greater than or equal to 20 years in the Union had increased risk of mortality from non-melanoma skin cancer with a standardized mortality ratio (SMR) of 4.0 that was statistically significant with a 95% confidence interval on the SMR of 1.0 to 10.9. USEPA (2013) has misreported the findings of Hammond et al. (1976). The study, itself, does not report confidence intervals, but IARC (2013) presents confidence intervals from the study. The 95% confidence interval on the SMR spans 0.82 to 11.69. Thus, the reported increase in skin cancer mortality is *not* statistically significant because of the small numbers. The number of cases was reported as only 3. Hammond et al. (1976) also report that their control group, the entire male population of the US, was “not ideal.”

Next, three studies are presented and classified as studies of workers exposed to creosote. Creosote is a complex mixture that contains low levels of BaP and other high molecular weight PAHs that have USEPA promulgated or proposed Relative Potency Factors. The major constituents of creosote are low molecular weight PAHs with no Relative Potency Factors. These three studies, discussed below, provide little credible evidence that BaP causes skin cancer in humans.

Pukkala (1995) performed a survey of 109,000 cancer cases and found 5 cases of nonmelanoma skin cancer in a group of “round-timber workers.” The classification of occupations was based on the Nordic Classification of Occupations of 1963. In this classification, the category 67 – Woodwork was subcategorized into 9 groups as follows:

- 670 Round-timber workers

- 671 Timber workers
- 672 Plywood makers
- 673 Construction carpenters
- 674 Boat builders
- 675 Bench carpenters
- 676 Cabinet makers
- 677 Woodworking machine operators, etc.
- 678 Wooden surface finishers
- 679 Woodworkers NOS (*other woodwork*)

It is clear from the classification scheme that these are not workers in creosote plants. The Farlex dictionary defined round timber simply as “felled trees which have not been converted to lumber.” In Finland, the word roundwood is simply used to refer to felled trees. The Finnish Statistical Yearbook of Forestry (Finnish Forest Research Institute, 2003) uses the word “roundwood” to discuss felled timber. This study is not a study of creosote workers. Appendix A of Pukkala (1995) also lists the total number of “round-timber workers” as 1,298. There may or may not be any wood treatment facilities in Finland, but it is inconceivable that, if there were, there would be enough facilities to keep 1,298 people employed. In conclusion, the Pukkala (1995) study has nothing to do with BaP.

It is not stated that these workers were working in wood treating facilities where wood is impregnated with creosote or other wood preserving agents. They are just as likely to be timber or sawmill workers. USEPA (2013) correctly reports that the 95% confidence interval on the SIR of 4.62 is (1.51-10.8), but the study is irrelevant to drawing conclusions about creosote or BaP. In addition, Pukkala et al. (2009) did not include this category of “round-timber workers” in their followup study.

The statistics are reported correctly from Karlehagen et al. (1992), but USEPA fails to note that the study did not control for exposure to sunlight and that the authors did not conclude that there was any correlation between creosote exposure and skin cancer.

These researchers studied the cancer incidence data on 922 timber creosoters at 13 plants in Sweden and Norway. Most cancer rates were not elevated compared to national statistics. Specifically, there was no increase in lung cancer, bladder cancer, or other cancers. However, lip cancer and nonmelanoma skin cancer rates were elevated. Of these, the lip cancer rate was not statistically significantly elevated compared to national statistics and can be discounted. The nonmelanoma skin cancer rate was statistically significantly elevated, and the rate was 2.4 times higher than national rates. This increase has been cited by some,

including USEPA, as definitive evidence that creosote does increase the risk of skin cancer in humans. However, the study has several significant methodological flaws that must be considered when evaluating the significance of the study.

First, the reported increase is due only to five excess cases compared to the expected number of cases in a group of 922 people. More importantly, the study did not control for exposure to sunlight. According to the authors, the excess skin cancers “could probably be attributed to the combination of exposure to creosote and sunlight.” This is due partly because the exposed workers had greater contact to sunlight than did the control population (national cancer rates) to which worker cancer rates were compared. Specifically, the authors stated: “as to the difference in cancer rates between urban and rural areas, the use of national rates could well have introduced bias because most of the plants were located in rural areas.” Thus, the five cases of skin cancer noted in the study were likely caused by ultraviolet light, which is the major cause of skin cancer humans.

When discussing their results in 1991 before the paper was published, the authors stated: “This study does not confirm that exposure to creosote in the wood preserving industry has caused an excess of total cancer morbidity. The study indicates, however, that exposure to creosote might increase the incidence of skin cancer.” Thus, the authors did not conclude that there was any correlation between creosote exposure and skin cancer.

USEPA (2013) also cites Tornqvist et al. (1986), but this study is not relevant to BaP exposures. These authors studied cancer in the electric power industry to investigate the hypothesis that occupational exposure to electromagnetic fields might be associated in increased cancer rates. Power linemen are clearly exposed to more sunlight than the normal population, yet there is no mention of the potential effects of sunlight. It is a mystery why USEPA (2013) reports a non-significant increase in non-melanoma skin cancer in workers exposed to sunlight with no reported risk factors for exposure to BaP as supporting evidence for their hypotheses. USEPA (2013) does correctly report that this paper shows no significant increase in skin cancer in power linemen.

USEPA (2013) correctly reports that Roelofzen et al. (2010), Pittlekow et al. (1981) and Maughan et al. (1980) showed no increased in skin cancer in pharmaceutical coal tar users, but they failed to present several other reports on pharmaceutical coal tar users as discussed below. USEPA (2013) also report findings from Stern et al. (1980, 1998) which are irrelevant because the population was a population of patients exposed to carcinogenic psoralens. The references to these two studies should be removed.

The following table summarizes the reports cited by USEPA and provides the actual reported statistical significance level of any conclusions if statistical testing was performed.

Summary of Reports Cited by USEPA (2013) Alleging Human Skin Cancer

Citation	Worker Group	Studied Effect Reported by USEPA (2013)	Statistical Significance Reported by USEPA (2013)	Actual Reported Statistical Significance
Spinelli et al. (2006)	Aluminum	Melanoma	Not significant	Not significant
Gibbs et al. (2007a,b)	Aluminum	Melanoma	Not significant	Not significant
Brown and Thornton (1957)	Chimney Sweeps	Scrotal cancer	Not reported & irrelevant	Not reported & irrelevant
Hammond et al. (1976)	Roofers/Water-proofers	Non melanoma skin cancer	Significant	Not significant
Pukkala (1995)	Round-timber workers	Non melanoma skin cancer	Significant	Not relevant*
Karlehagen et al. (1992)	Creosote wood treatment workers	Non melanoma skin cancer	Significant	Not relevant*
Tornquist et al. (1986)	Power linesmen	Non melanoma skin cancer	Not significant	Not significant
Roelofzen et al. (2010)	Coal Tar Pharmaceutical Users	Non melanoma skin cancer	Not significant	Not significant
Pittlekow et al. (1981)	Coal Tar Pharmaceutical Users	Non melanoma skin cancer	Not significant	Not significant
Maughan et al. (1980)	Coal Tar Pharmaceutical Users	Non melanoma skin cancer	Not significant	Not significant
Stern et al. (1998)	Patients Exposed to Carcinogenic Psoralens	Non melanoma skin cancer	Not reported	Not relevant*
Stern et al. (1980)	Patients Exposed to Carcinogenic Psoralens	Non melanoma skin cancer	Not reported	Not relevant*

* See text

In sum, there is no credible evidence reported in the above studies that links BaP exposure to human skin cancer.

USEPA (2013) further cited IARC reports that served as the basis for their classifications that certain mixtures or occupations are associated with increases in skin cancer and may have some relationship to BaP exposures.

The studies mentioned in USEPA's Table 1-10 are all IARC documents. Baan et al. (2009), Benbrahim-Tallaa et al. (2012) and Secretan et al. (2009) are journal summarizations of recent IARC Monographs. They are not primary scientific articles.

The following mixtures or occupations are listed as having "sufficient evidence in humans" or "limited evidence in humans" of skin cancer.

**Summary of Mixtures/Occupations Cited by USEPA (2013)
As Evidence of Skin Cancer in Humans**

Mixture or Occupation	IARC Classification Sufficient	IARC Classification Limited
Coal Tar Distillation	X	
Creosotes		X
Mineral oils, untreated or mildly treated	X	
Shale Oils	X	
Soot (chimney sweeping)	X	

Coal Tar Distillation: Baan et al. (2009) is cited by USEPA (2013) in Table 1-10, but this article presents no primary sources. To find the basis for this classification, one needs to consult IARC (2012b) Monograph 100F. The literature cited is listed below as the basis for the “sufficient” classification.

- Henry (1946)
- Letzel & Drexler (1998)

Neither of these studies was an epidemiological study. They were health surveys. In both cases, they describe the skin lesions in workers exposed to extremely high levels of multiple chemicals over long periods of time during the early part of the twentieth century when worker hygiene practices were nonexistent.

IARC (2010b) correctly reports that Henry (1946, 1947) lists skin cancers (cutaneous epitheliomata) from many trades, such as coal tar distilling, as well as cotton manufacture, patent fuel manufacture, coal gasification, coke oven workers, shale oil refining, and others. For 1920-1945, the number of tumors for tar distillers was listed as 939, but the number of people was listed as 538. Of the lesions, 81% were on the head, neck and upper limbs, and 17% were on the scrotum. The tumors noted in the 538 workers were attributed to “pitch, tar, and tar-products, not solely to “coal tar” as noted in IARC (2010b). It is important to note that the study gives no exposure information, no information on co-exposures, no information on type of tar used in the distillery and no information on any control population.

IARC (2010b) correctly reports that 151 workers were listed as having squamous cell carcinoma and 98 with basal cell carcinoma in a German tar distillery from 1946 to 1996 (Letzel & Drexler, 1998). The study was a retrospective survey. Data were collected retrospectively from insurance records. Workers birth dates varied from 1882 to 1960 with a median of 1922. 204 of the workers had started working by 1940. Thus, many of the cases were due to exposures received early in the century. More than 77% of the workers were employed for more than 20 years (median employment was 32 years). It is important to note that the study gives no exposure information, no information on co-exposures, no information on type of tar used in the distillery and no information on any control population.

USEPA (2013) has mis-cited Baan et al. (2009) and IARC (2010b) by summarizing this classification as evidence that “skin cancers are the cancers that have been observed in occupational studies of PAH mixtures.” The above two citations were surveys of diseases reported in workers in a specific industry in the early days of the last century when worker hygiene practices were negligible. The commenter questions the relevance of IARC’s classification of “sufficient evidence in humans for the carcinogenicity of occupational

exposures during coal-tar distillation” [during the first half of the twentieth century] {noted added} to the allegation that BaP or BaP-containing mixtures at modern day exposure levels and with modern hygiene practices have been shown to have caused cancer in any humans.

Furthermore, IARC (1985, 2010b) does not classify coal tar as having sufficient or even limited evidence that it causes cancer in humans. IARC has been careful to only list occupational exposures during coal tar distillation as Group 1. (<http://monographs.iarc.fr/ENG/Classification/ClassificationsAlphaOrder.pdf>).

Creosotes: In Table 1-10, USEPA (2013) also stated that the IARC classification that there is limited evidence in humans that creosotes have caused cancer is evidence that BaP causes cancer in human skin. The current IARC classification is cited in IARC (1987) as follows: “In a number of case reports, the development of skin cancer in workers exposed to creosote is described. One study involved a review of 3,753 cases of cutaneous epithelioma from 1920 to 1945 and showed that 35 cases (12 of which were of the scrotum) had had exposure to creosote. Most cases occurred in workers handling creosotes or creosoted wood during timber treatment. A mortality analysis of workers in many occupations indicated an increased risk of scrotal cancer for creosote-exposed brickmakers [ref: 1]. “

Reference 1 is IARC (1985) and this document provides only five citations that provide any information on the potential for creosote oils to cause skin cancer:

- O’Donovan (1920)
- Cookson (1924)
- Henry (1947)
- Henry (1946)

O’Donovan (1920) is correctly identified by IARC (1985) as a “case report.” The author reported carcinomas in 4 heavily exposed creosote workers from 1903-1920 who were exposed for > 20 years (20, 30, > 30 years, and “always”). The major commonality among these workers was “the long duration of employment that is common before the development of serious lesions.” According to the author, other skin lesions were present for many years before the tumor developed. The author stated: “in this industry, generally, employment is fairly stable; cases of continuous employment from boyhood to old age are not rare, and hence continuous exposure is the lot of many workers.”

Cookson (1924) is another case report. One worker who was exposed for 33 years exposure to creosote in a creosote factory, carrying creosoted lumber was reported to have a skin cancer. According to the author, the worker was “up to his eyes in creosote,” indicating that he was heavily exposed for 33 years.

Henry (1947) reported 35 cases of skin cancer among 34 workers during the years, 1920-1945. No information is provided about co-exposures, exposure conditions, or other important factors. Henry (1947)

merely reported that about one person per year in the United Kingdom during 1920-1945 presented with skin cancer after presumably high level exposure to creosote over a long period of time.

IARC (1985) lists 9 cases of scrotal cancer from Henry (1946) in brickmakers allegedly exposed to creosote oil from 1911-1938, but 8 of those cases were reported in Henry (1947) when he summarized the skin cancer data of brickmakers from 1920-1945. Thus, this report presents one case not summarized in Henry (1947).

Thus, the sum total of evidence that coal tar creosote may have caused skin cancer in humans is a total of 40 case reports over the period 1920-2010, when IARC reviewed the classification. That's less than one case per year. Clearly, the world literature upon which the classification is based is, indeed, limited.

Furthermore, neither IARC nor USEPA (2013) has cited the well-designed and clearly relevant epidemiology study of Wong and Harris (2005). These authors performed a nested case-control study of creosote workers from eleven wood treating plants in the United States. The study involved 2,179 workers at 11 plants where railroad ties and utility poles were treated with coal-tar creosote. The period of observation was 1979-2001. The authors' conclusion was: "Based on the present investigation and other studies, there was no evidence that employment at the 11 wood-treating plants or exposure to creosote-based wood preservatives was associated with any significant mortality increase from site-specific cancers or nonmalignant diseases." Melanoma skin cancer was not increased compared to expected levels."

Mineral oils, untreated or mildly treated: Baan et al. (2009) is cited by USEPA (2013) in Table 1-10, but this article presents no primary sources. To find the basis for this classification, one needs to consult IARC (2012) Monograph 100F. IARC (2012b) has stated that mineral oils are a diverse group of mixtures and that it is very difficult to attribute the risk to specific chemical agents. IARC (2012b) does not attribute the carcinogenicity of mineral oils to BaP.

Shale oils: Baan, et al. (2009) is cited by USEPA (2013) in Table 1-10, but this article presents no primary sources. To find the basis for this classification, one needs to consult IARC (2012b) Monograph 100F. IARC (2012b) has stated that "Crude shale oils are very complex mixtures, and only a few compounds have been identified." IARC (2012b) does not attribute the carcinogenicity of shale oils to BaP.

Soot (chimney sweeping): Baan, et al. (2009) is cited by USEPA (2013) in Table 1-10, but this article presents no primary sources. To find the basis for this classification, one needs to consult IARC (2012b) Monograph 100F. IARC (2012b) defined the mixture specifically as "soot, as found in occupational exposure of chimney sweeps." The literature cited is listed below as the basis for the "sufficient" classification.

- Pott (1775)
- Earle (1808)
- Butlin (1892)

- Henry and Irvine (1936)
- Henry (1937)
- Henry (1946)
- Henry (1947)
- Evanoff et al. (1993)
- Pukkala (1995)

As mentioned above, Pott (1775) was a medical lecture that presented general information about scrotal cancer in English chimney sweeps. Earle (1808) was a similar case report. IARC (2012b) has cited the work of Butlin (1892) *Cancer of the Scrotum in Chimney Sweeps and Others, Lecture II, Why Foreign Sweeps Do Not Suffer from Scrotal Cancer*. In this lecture, Butlin (1892) states that “sweep’s cancer of the scrotum is a disease which is almost unknown in the large European countries and in the United States of America.” His research pointed to differences in coal used in Britain and other countries as well cleanliness of clothing and frequency of bathing. The commenter is not disputing that scrotal cancer occurred in chimney sweeps in the 18th and 19th centuries, but rather that these historical reports have little relevance to making a modern day determination about the role of BaP in current human skin cancers due to the extremely large doses that these early chimney sweeps received, which have no relevance to modern day exposures.

Henry (1937) does report that 7% of the cases of scrotal cancer in England and Wales from 1911-1985 were in chimney sweeps, but Henry (1946) and Henry (1947) do not report any skin cancer cases for chimney sweeps, and “soot” is not listed as a causal agent in any of the reported cases. Again, these data from Henry (1937) refer just to British chimney sweeps, and no data are in the literature for chimney sweeps or others exposed to soot from any other country, including the United States.

IARC (2012b) also mentions an epidemiology study of Swedish chimney sweeps who worked from 1917 to 1980. Evanoff, et al. (1993) specifically evaluated nonmelanoma skin cancer and found no increase in skin cancer compared to expected levels in Swedish chimney sweeps. The most recent update to this cohort (Hogstedt et al., 2013) also shows no increase in melanoma and non melanoma skin cancer in the cohort of chimney sweeps.

Pukkala et al. (2009) specifically followed 5,498 male chimney sweeps in Denmark, Finland, Norway and Sweden and found no excess of nonmelanoma skin cancer. According to IARC (2012b): “Despite the classical risk for scrotal cancer in chimney sweeps, studies of this occupational group under modern working conditions shows no such excess.”

IARC (2012b) specifically states that their classification of soot in chimney sweeps is based on history. Specifically: “From historical case reports there is sufficient evidence of an increased risk for (scrotal) skin cancer among chimney sweeps.” What IARC (2012b) *should have concluded* is that from historical case

reports there was sufficient evidence of an increased risk for (scrotal) skin cancer among chimney sweeps in the past, but this is no longer the case under modern working conditions and hygiene practices.

Summary of Discussion on Skin Cancer in Humans: In conclusion, USEPA (2013) has presented no human studies that link exposure to BaP with human skin cancer. The human studies that have been presented are studies of worker groups who were exposed to complex mixtures. It is not possible to attribute the effects seen in complex mixtures to one compound that may be present at only trace levels in the mixture.

In addition, USEPA (2010) proposed that benzo(c)fluorene, benz[j]aceanthrylene, benz[l]aceanthrylene, dibenz[a,h]anthracene, and dibenzo[a,l]pyrene were all more potent than BaP in causing cancer, mostly based on mouse skin studies. Thus, BaP may not be the causal agent in the mixtures cited by USEPA. Instead, other PAHs or even non-PAH compounds may be the causal agents in any effects that might be seen after exposures to the mixtures.

More importantly, even if one assumes for the moment that data on coal tar distillation, creosote, and soot exposure to chimney sweeps is somehow relevant to evaluating BaP, the human data presented by USEPA (2013) and IARC are extremely limited, old, and relevant only to industrial work conditions and hygiene practices that no longer exist. There are no modern day epidemiology studies that demonstrate that BaP is a cause of human skin cancer in today's world.

Furthermore, the Food and Drug Administration (FDA, 2001) concluded: "Upon reviewing the published studies, the agency does not find that there is evidence to implicate the use of OTC coal tar containing drug products as an independent risk factor for the development of skin cancer."

Lung Cancer in Humans

USEPA (2013) has mischaracterized the weight of evidence that BaP causes lung cancer in humans.

On page 1-51, the document states three studies, labeled "Tier 1" studies observed increased risk of lung cancer with increasing exposure to BaP "although the relative contributions of benzo[a]pyrene and of other PAHs cannot be established." They are erroneously cited in Table 1-11 as "studies of benzo(a)pyrene" which they are not. These three studies are not studies of BaP; they are studies of workers in industries having exposures to complex mixtures. Each is discussed below.

- Armstrong and Gibbs (2009)
- Spinelli et al. (2006)
- Xu et al. (1996)

Armstrong and Gibbs (2009) was a follow up study of 16,431 workers who had worked in aluminum smelters in Quebec from 1950 to 1999. Aluminum smelters using the Soderberg process use a specific PAH-containing mixtures, petroleum coke and coal tar pitch, as the electrodes in a very high temperature (1,000 degrees Centigrade) process to make aluminum.

In this large group of workers, 514 lung cancers were expected and 677 were observed. The SMR was statistically significantly increased, but the lung cancers cannot be attributed to BaP exposure from the aluminum manufacturing process. Many of these cancers were, in fact, attributed to cigarette smoking. The authors conclude: "...the shape of the exposure-response function and the mode of combination of risks due to occupational PAH and smoking remains uncertain."

Spinelli et al. (2006) is also a study of 6,423 aluminum plant workers. In this study, a Job Exposure Matrix was developed using benzene soluble material and BaP measurements. A total of 2624 personal benzene soluble materials measurements and 1275 personal BaP measurements were taken from 1975-2001. The data are not useful for making any statements about BaP levels because the samples were collected on fiberglass filters and then desorbed with benzene. BaP boils at 495 degrees Centigrade, and the aluminum processes run at 1,000 degrees Centigrade, so sampling cassettes were only capturing the particle-bound BaP, not the vapor phase BaP. A sampling approach that only captures particle-phase BaP is perhaps adequate for some workers, but not potroom workers who are working in areas where the temperatures are extremely high.

These authors found that neither cancer incidence nor cancer mortality was statistically significantly increased for lung cancer or cancer of the oropharynx, nasopharynx, hypopharynx, nose, or larynx. Despite the fact that there was no increased lung cancer risk overall in the facilities, a trend of increasing lung cancer incidence was reported with increasing doses of particulate BaP, and the trend was statistically significant. It is important to note that the only dose category showing a statistically significant increase over the expected numbers of lung cancers was the $>80 \mu\text{g}/\text{m}^3\text{-years}$ group. This reported "partial" cumulative dose underestimates the true dose to which workers were exposed because the vapor phase BaP was ignored.

Xu et al. (1996) was a nested case-control study of lung cancer in among iron and steel workers. The number of lung cancer cases among active and retired employees of the Anshan iron-steel complex was 610. It is stated that there were 82,867 historical monitoring records for dust and benzo(a)pyrene from 1956 to 1992. The paper does not report how many of these measurements were dust measurements and how many were BaP measurements. These commenters find it implausible that industrial hygienists collected thousands of BaP measurements over this period of time. In any case, no information is presented regarding the method of sample collection or analysis for BaP. The study also reports: "Since results were similar for average total dust and BaP, large particle total dust, and small particle total dust, only results for cumulative total dust and BaP are presented here." It is impossible for the concentration of total dust to equal the concentration of BaP in any work place at any time.

The odds ratio for lung cancer was reported to be statistically significantly elevated in several groups, some of which were said to have BaP exposure and some of which were said to have not had BaP exposure. These findings are not strong, because the manner in which the odds ratios were calculated are unclear.

Table III reports cases and controls and the total number of controls is 643. However, the footnote states: “All OR relative to risk for the 172 case and 411 controls subjects who were employed only in administrative or low-exposure occupations (nonexposed).” The sum of 172 and 411 is 583, so how the odds ratios were calculated are suspect, and any reported increase in odds ratio is uninformative if the ratios were calculated incorrectly. In addition, as reported below, the author’s state that the study used 1,100 controls; however, no facts about the number and use of controls are consistent throughout the paper.

95% Confidence Intervals of Odds Ratios in Xu et al. (1966)

Occupation Groups	Ever Worked	Worked >15 years	PAH Exposure?	Total # of cases out of 610
Smelting & rolling	1.2-2.1*	1.1-2.2*	+ (smelting only)	166 ¹
Foundry worker	1.1-2.8*	0.8-2.4**	+	48
Fire-resistant brick	1.0-2.9***	1.4-5.9*	no	43
Coke oven worker	1.7-7.5*	1.4-8.5*	+++	22
Total				279

* Significant; ** Not significant; *** Marginally significant

¹ The paper does not state if these lung cancer cases were in smelter workers or roller workers.

In addition to the odds ratios for specific occupational groups, Xu et al. (1996) (see Table IV) calculated odds ratios for cases and controls with regard to estimated cumulative BaP exposure, but these odds ratios are seriously flawed. Regardless of the occupational group, the authors somehow assigned 390 lung cancer cases to one of four BaP dose groups: <0.84, 0.85-1.96, 1.97-3.2 and ≥ 30 ($\mu\text{g}/\text{m}^3$ -year). First, it makes no sense to leave out a group having an exposure of 3.3-30 $\mu\text{g}/\text{m}^3$ -year. This discontinuity is inappropriate.

More importantly, the use of controls in the odds ratios is totally illogical. Odds ratios in case-control studies compare cases (in this case lung cancer cases) with matched controls that have the same characteristics as the cases, but without the potential exposures under study. Xu et al. (1996) describe the selection of controls as follows:

A computerized personnel file, created in 1989, was used to randomly select controls from the 196,993 active and retired employees of the iron-steel complex. The file was sorted by birth date, and controls were group matched by gender and birth year according to the age and gender distributions of these cancers during the time period 1980- 1989. A total of 1,100 controls, each of whom had worked for at least 10 years at the complex, were selected for interview.

This sounds reasonable, but Table IV compares cases in four BaP exposure groups to controls for those same groups. Xu et al. (1996) do not explain how the 114 controls for the low BaP dose group were selected or how the 117 controls for the high BaP dose group were selected. To further confuse the issue, the author's state: "All OR relative to risk for the 172 case and 411 controls subjects who were employed only in administrative or low-exposure occupations (non-exposed). Without thorough understanding of who the controls were, the odds ratios are meaningless.

Furthermore, Table IV shows that the risk of lung cancer is associated with all three exposure metrics: total dust, silica dust and BaP. In fact, the risk of lung cancer is reported to be more associated with total dust than it is with BaP. Accordingly, no information is presented in this paper about the role of BaP in the reported lung cancers in workers at this iron and steel complex.

Summary of Tier 1 Reports Cited by USEPA (2013) Alleging Human Lung Cancer

Citation	Worker Group	Studied Effect Reported by USEPA (2013)	Statistical Significance Reported by USEPA (2013)	Actual Reported Statistical Significance
Armstrong and Gibbs (2009)	Aluminum workers	1. Increase in lung cancer 2. Lung cancer increased with increasing BaP exposure levels	1. Significant 2. Significant	1. Significant 2. Significant
Spinelli et al. (2006)	Aluminum workers	1. Increase in lung cancer 2. Lung cancer increased with increasing BaP exposure levels	1. Significant 2. Significant	1. Not significant 2. No association with BaP possible, because only particulate BaP measurements were made
Xu et al. (1996)	Iron and steel workers	1. Increase in lung cancer 2. Lung cancer increased with increasing BaP exposure levels	1. Significant 2. Significant	1. Significant, but unexplained manner of calculating odds ratios makes all claims suspect 2. No association with BaP possible, because no information provided about collection and analysis of BaP samples

USEPA (2013) also cites Tier 2 studies in Table 1-12. They are erroneously cited in Table 1-12 as “studies of benzo(a)pyrene” which they are not. These three studies are not studies of BaP; they are studies of workers in industries having exposures to complex mixtures. USEPA (2013) also mischaracterizes the studies by stating: “This discussion primarily focuses on epidemiologic studies that included a direct measure of benzo[a]pyrene exposure.” Few of these studies made direct measurements of BaP in the work environment. Each is discussed below.

- Friesen et al. (2009)
- Olsson et al. (2010)
- Costantino et al. (1995)
- Liu et al (1997)
- Berger and Manz (1992)
- Hansen (1989,1991)
- Gustavsson et al. (1990) [actually Gustavsson & Reuterwall (1990)]
- Moulin et al. (1989)
- Hammond et al. (1976)

Friesen et al. (2009) showed no statistically significant increase in lung cancer in 4,316 workers who were employed for more than 90 days between 1983 and 2002 at two aluminum plants in Australia. There was no statistically significant trend in relative risk for lung cancer with increasing cumulative BaP exposure. BaP exposure was estimated using a job exposure matrix based on job classifications and previously collected BaP measurements using unreported sample collection and analysis methods.

Olsson et al. (2010) studied 433 lung cancer cases and 1,253 controls in a study of European asphalt workers exposed to bitumen, asbestos, coal tar, crystalline silica, and diesel exhaust. BaP was not measured or estimated. According to the authors, semiquantitative exposure estimates were obtained for 4- to 6-ring PAHs from an industry Asphalt Workers Exposure Database. No statistically significant increases in the odds ratio for lung cancer were found for workers exposed to bitumen fume, PAH or bitumen condensate. No statistically significant increases in the odds ratio for lung cancer were found for workers exposed to high levels of bitumen fume (inhalation) or bitumen condensate (dermal) as measured by years, exposure units, or exposure unit-years. No statistically significant trends were seen as exposure increased. No statistically significant increases in the odds ratio for lung cancer were found for workers exposed to coal tar as measured by years or exposure units. No statistically significant trends were seen as exposure increased. The only statistically significant result found was in this highest cumulative coal tar exposure

group when exposure was assessed by unit-years. Three of four groups had no statistically significant increase in odds ratio, but the highest group had a small statistically significant result, but the trend analysis did not show a statistically significant increase in odds ratio as unit-years increased. In summary, Olsson et al. (2010) presented 49 comparisons of odds ratios in this study. Forty-eight of the comparisons were not statistically significant, and one was.

Costantino et al. (1995) reported a statistically significant increase in lung cancer deaths in 5,321 coke oven workers. The relative risk increased with increasing number of years as a coke-oven worker. BaP was not measured or estimated. Only coal tar pitch volatiles were measured.

Liu et al. (1997) reported a statistically significant increase in lung cancer in 6,635 carbon electrode workers from six carbon electrode companies and one aluminum smelter employed for more than 15 years. IARC noted that "this study included an unspecified number of aluminum reduction plant workers in addition to carbon electrode manufacturing workers." When the workers were grouped by no, low, moderate and high contact with carbon compounds, the increase lung cancer mortality was significant only in the highest exposure group. No historical PAH exposure information was available, and exposure classifications were not done on the basis of any chemical concentration data. The groups were defined in an *ad hoc* manner based on the types of processes involved. Benzene soluble compounds and BaP were measured after the fact in one of the seven plants. There were 11 high exposure group samples, 3 moderate exposure group samples, and 1 plant site sample. BaP concentrations were higher in the high exposure group, but the BaP data were not used in any way in the study. USEPA (2013) has incorrectly reported BaP concentrations for the "moderate" and "high" exposure groups from the seven facilities over time by inserting the BaP data from a limited number of measurements in one plant in one year into a table, misleading the reader into thinking that those BaP measurements were linked to the SMRs from those worker groups. This is entirely false.

Berger and Manz (1992) found a statistically significantly increased rate of lung cancer mortality in 789 coke oven workers who worked for more than 10 years in a gas plant and began working from 1900-1989. BaP was measured in the plant in 1983, but no trend analysis was performed to determine whether lung cancer risk increased with increasing exposure to BaP and BaP exposure levels were not assigned to any groups or individuals.

Hansen (1989, 1991) found a statistically significantly increased rate of lung cancer mortality in 679 asphalt workers employed from 1959-1980. Asphalt fume condensate was measured in 35 personal samples. The dates, locations, collection method and analytical method were not reported. BaP was determined in 9 of these samples. Again, no information is presented regarding the manner in which the samples were collected or analyzed. No trend analysis of risk versus BaP level was performed.

Gustavsson et al. (1990) is a mis-citation in USEPA (2013) Table 1-12. The data presented is from Gustavsson & Reuterwall (1990). In this study, 295 gas production workers were studied, and no significant increase in lung cancer incidence or mortality was seen in the entire cohort or just those working at the coke ovens. BAP was measured in 1964 and 1965 but no trend analysis was performed and the workers were not categorized by their exposure level to BaP or any other chemical.

Moulin et al. (1989) found no statistically significant increase in lung cancer incidence in 1,302 carbon electrode workers in one plant or 1,115 carbon electrode workers in a second plant. No historical exposure information on BaP was available. Particulate BaP was measured at the time of the study in an unknown number of samples. No trend analysis of risk versus BaP level was performed.

Hammond et al. (1976) found an increased number of lung cancer deaths compared to expected rates for pavers and roofers exposed for 20 or more years (22 deaths out of 4,215 workers), but not in workers exposed for 9-19 years (99 deaths out of 1,724 workers). No statistical significance testing was done. Particulate BaP was measured by an unreported method in 52 samples from 8 roofing job types, but no trend analysis was performed.

Summary of Tier 2 Reports Cited by USEPA (2013) Alleging Human Lung Cancer

Citation	Worker Group	Studied Effect Alleged by USEPA (2013)	Statistical Significance Reported by USEPA (2013)	Actual Reported Statistical Significance
Friesen et al. (2009)	Aluminum workers	1. Increase in lung cancer 2. Lung cancer increased with increasing BaP exposure levels	1. Not Significant 2. Not Significant	1. Not Significant 2. Not Significant. BaP exposures estimated from job exposure matrix and BaP measurements using unreported methods.
Olsson et al. (2010)	Asphalt workers	1. Increase in lung cancer 2. Lung cancer increased with increasing BaP exposure levels	1. One significant & 3 nonsignificant results 2. One nonsignificant trend with coal tar	1. One significant & 39 nonsignificant results 2. Nine nonsignificant trends. BaP not measured or estimated. PAH levels estimated from external database.
Costantino et al. (1995)	Coke oven workers	1. Increase in lung cancer 2. Lung cancer increased with increasing BaP exposure levels	1. Significant 2. Significant	1. Significant 2. No association with BaP possible, because only Coal Tar Pitch Volatiles measured.
Liu et al (1997)	Carbon electrode manufacture	1. Increase in lung cancer 2. Lung cancer increased with increasing BaP exposure levels	1. Significant 2. Significant	1. Significant 2. No trend data was presented; BaP data presented by USEPA is from one plant & onetime point and not associated with the SMR data.

Citation	Worker Group	Studied Effect Alleged by USEPA (2013)	Statistical Significance Reported by USEPA (2013)	Actual Reported Statistical Significance
Berger and Manz (1992)	Coke oven workers	1. Increase in lung cancer 2. Lung cancer increased with increasing BaP exposure levels	1. Significant 2. No trend information presented	1. Significant 2. No trend information presented; BaP measured 10 years earlier but not used in study.
Hansen (1989,1991)	Asphalt workers	1. Increase in lung cancer 2. Lung cancer increased with increasing BaP exposure levels	1. Significant 2. No trend information presented	1. Significant 2. No trend information presented
Gustavsson & Reuterwall (1990)	Coke oven workers	1. Increase in lung cancer 2. Lung cancer increased with increasing BaP exposure levels	1. Not Significant 2. No trend information presented	1. Not Significant 2. No trend information presented
Moulin et al. (1989)	Carbon electrode workers	1. Increase in lung cancer 2. Lung cancer increased with increasing BaP exposure levels	1. Not Significant 2. No trend information presented	1. Not Significant 2. No trend information presented; No historical BaP information was available.
Hammond et al. (1976)	Paving workers	1. Increase in lung cancer 2. Lung cancer increased with increasing BaP exposure levels	1. Significant 2. No trend information presented	1. No significance testing performed 2. No trend information presented

In sum, there is little credible evidence reported in the above studies that links BaP exposure to human lung cancer. Workers in high temperature environments, such as aluminum production workers and coke oven workers, have been shown to have increased rates of lung cancer compared to the general population, but the role of BaP versus the other hundreds of chemicals to which these workers are exposed is unexplained.

USEPA (2013) further cited IARC reports that served as the basis for their classifications that certain mixtures or occupations are associated with increases in lung cancer and may have some relationship to BaP exposures.

The studies mentioned in USEPA's Table 1-10 are all IARC documents. Baan et al. (2009), Benbrahim-Tallaa et al. (2012) and Secretan et al. (2009) are journal summarizations of recent IARC Monographs. They are not primary scientific articles.

The following mixtures or occupations are listed in USEPA (2013) as having “sufficient evidence in humans” or “limited evidence in humans” of lung cancer.

**Summary of Mixtures/Occupations Cited by USEPA (2013)
As Evidence of Lung Cancer in Humans**

Mixture or Occupation	IARC Classification Sufficient	IARC Classification Limited
Aluminum production	X	
Carbon electrode manufacture		X
Coal gasification	X	
Coal tar pitch (paving and roofing)	X	
Coke production	X	
Diesel exhaust	X	
Indoor emissions from household combustion of biomass fuel (primarily wood)		X
Indoor emissions from household combustion of coal	X	
Soot (chimney sweeping)	X	

Aluminum Production: Baan et al. (2009) is cited by USEPA (2013) in Table 1-10, but this article presents no primary sources. The most recent IARC discussion of aluminum production is found in IARC (2012b) Chemical Agents and Related Occupations (Volume 100 F). The literature cited is listed below as the basis for the “sufficient” classification.

- Gibbs et al. (2007)
- Gibbs and Sevigny (2007a,b)
- Armstrong and Gibbs (2009)
- Bjor et al. (2008)
- Spinelli et al. (2006)
- Friesen et al. (2007)
- Friesen et al. (2009)
- Sim et al. (2009)
- Mur et al. (1987) 5404
- Moulin et al. (2000)

- Romundstad et al. (2000)
- Rockett and Arena (1983) 5497

Gibbs et al. (2007) is an earlier report from the same cohort as reported in Armstrong and Gibbs (2009). It is not an independent source of information on lung cancer in aluminum workers.

Gibbs and Sevigny (2007a) is an earlier report from the same cohort as reported in Armstrong and Gibbs (2009). It is not an independent source of information on lung cancer in aluminum workers.

Armstrong and Gibbs (2009) was discussed above. There was a statistically significantly increased risk of cancer in aluminum workers.

Bjor et al. (2008) found a statistically significant increase in lung cancer in 2,264 workers employed for more than 0.5 years from 1942-1987 in an aluminum plant in Sweden. There was no statistically significant increase in workers there for less than 10 years, but it was significant in workers there for more than 10 years.

Spinelli et al. (2006) was discussed above. These authors found that neither cancer incidence nor cancer mortality was statistically significantly increased for lung cancer or cancer of the oropharynx, nasopharynx, hypopharynx, nose, or larynx.

Friesen et al. (2007) is a second report on the same workers and provides no independent information.

Friesen et al. (2009) showed no statistically significant increase in lung cancer in 4,316 workers who were employed for more than 90 days between 1983 and 2002 at two aluminum plants in Australia.

Sim et al. (2009) showed no statistically significant increase in lung cancer in the same cohort as Friesen et al. (2009).

Mur et al. (1987) showed no statistically significant increase in lung cancer in 6,455 aluminum workers who worked more than one year in one of 11 plants between 1950 and 1976.

Moulin et al. (2000) showed no statistically significant increase in lung cancer in 2,133 workers who worked more than one year between 1950 and 1994.

Romundstad et al. (2000) showed no statistically significant increase in lung cancer in 11,103 workers in 6 aluminum plants who worked more than 3 years between 1953 and 1996. There was no statistically significant increase in lung cancer even in workers estimated to have been highly exposed to PAHs.

Rockett and Arena (1983) showed no statistically significant increase in lung cancer in 21,829 workers in 14 plants who worked more than five years between 1946 and 1977.

Gibbs and Sevigny (2007b) is not cited in the IARC (2012b) text presumably because it found no association with estimated BaP exposures and because it found no statistically significantly increased risk of lung cancer in workers first hired from 1960-1969 or 1970-1979. A statistically significant increase in lung cancer was seen only in workers hired from 1950-1959. The authors conclude: "Cancer causes contributing to the large excess of cancer in the cohorts hired before 1950 (lung and bladder) seem to be diminishing."

Giovanazzi & D'Andrea (1981) is not cited in the IARC (2012b) text presumably because it found no statistically significant increase in lung cancer in 494 aluminum pot workers who worked from 1965-1979.

Carta et al. (2004) is also missing from the IARC text. It showed no increase in lung cancer in 1,152 men working for more than one year between 1972 and 1980 in an aluminum smelter.

Summary: IARC states: "An increased risk for lung cancer has been found in several but not all epidemiological studies in the aluminum-production industry. The exposure circumstances, especially levels of PAH in aluminum smelters, vary between industrial departments and also depend on the process used. However, data are not sufficient to disentangle the cancer risks associated with these different exposure situations."

Carbon electrode manufacture: IARC (2010a) (Volume 92) is cited by USEPA (2013) in Table 1-10. The literature cited is listed below as the basis for the "limited" classification.

- Teta et al. (1987)
- Moulin et al. (1989)
- Gustavsson et al. (1995)
- Liu et al. (1997)
- Donato et al. (2000)
- Mori (2002)
- Merlo et al. (2004)

Teta et al. (1987) found no statistically significant increase in lung cancer in 2,212 men employed at a carbon product department for at least 10 years.

Moulin et al. (1989) found no statistically significant increase in lung cancer in 1,302 and 1,115 workers in two carbon electrode plants.

Gustavsson et al. (1995) found no statistically significant increase in lung cancer in 901 workers at a graphite electrode plant employed for greater than 3 months from 1968 to 1988.

Liu et al. (1997) reported a statistically significant increase in lung cancer in 6,635 carbon electrode workers from six carbon electrode companies and one aluminum smelter employed for more than 15 years. IARC noted that “this study included an unspecified number of aluminum reduction plant workers in addition to carbon electrode manufacturing workers.” When the workers were grouped by no, low, moderate and high contact with carbon compounds, the increase was significant only in the highest contact group.

Donato et al. (2000) found no statistically significant increase in lung cancer in 1,006 graphite electrode workers employed for more than 1 year from 1945-1971.

Mori (2002) reported a statistically significant increase in lung cancer in 332 workers employed for more than 1 year at a graphite electrode manufacturing plant from 1951-1974.

Merlo et al. (2004) found no statistically significant increase in lung cancer in 1,291 workers at a graphite electrode manufacturing plant for more than 1 year from 1950-1989. When the cohort was broken down by number of years employed, there was no statistically significant increase even in those who worked for more than 19 years.

IARC (2010b) provided this summary which demonstrates that the human data on carbon electrode manufacturers is, indeed, limited at best:

“A study of carbon electrode workers in China showed an excess risk for lung cancer and a positive exposure–response relationship between increasing exposure to carbon compounds and lung cancer risk. When the study was limited to nonsmokers, the increased risk was still observed. However, the study included both carbon electrode workers and pot-room workers in an aluminum smelter, and it is questionable how much of the excess risk may be attributed to exposures in carbon electrode manufacture. A small study of carbon electrode manufacturing workers in Japan showed an excess incidence of lung cancer. A large study of workers at a carbon product department of a plant in the USA showed no excess incidence of respiratory cancer and no exposure–response trend in internal analyses. A cohort study of two plants in France and two cohort studies from Italy provided no evidence for an increased risk for lung cancer associated with carbon electrode manufacture. A small study from Sweden based on only two cases was uninformative due to small numbers.”

Coal gasification: Baan, et al. (2009) is cited by USEPA (2013) in Table 1-10, but this article presents no primary sources. The most recent IARC discussion of coal gasification is in IARC (2012b) *Chemical Agents and Related Occupations* (Volume 100 F). The literature cited is listed below as the basis for the “sufficient” classification.

- Doll et al. (1972)
- Berger & Manz (1992)

- Martin et al. (2000)
- Kennaway & Kennaway (1947)
- Kawai et al. (1967)
- Hansen et al. (1986)
- Wu (1988)

Doll et al. (1972) is considered by IARC (2010b) as a major source of information on coal gasification workers. IARC (2010) reports that lung cancer was statistically significantly increased in a subset of 1,176 coal gasification workers who had heavy exposure to coal gas. IARC (2010b) reports a relative risk of 179 with 95th confidence interval of 146-218 based on 99 cases. Doll et al. (1972) does not present relative risk ratios. It only reports the 99 cases of lung cancer in the 1,176 workers.

Berger and Manz (1992) found a statistically significantly increased rate of lung cancer mortality in 789 coke oven workers who worked for more than 10 years in a gas plant and began working from 1900-1989. BaP was measured in the plant in 1983, but no trend analysis was performed to determine whether lung cancer risk increased with increasing exposure to BaP and BaP exposure levels were not assigned to any groups or individuals.

Martin et al. (2000) is considered by IARC (2010b) as a major source of information on coal gasification workers. They reported a statistically significant increase in lung cancer deaths in coal gasification workers exposed at the highest of four categories, but there was no statistically significant increase in lung cancer risk at the other three exposure levels.

Kennaway & Kennaway (1947) found an excess in the number of cancer deaths for gas works laborers and gas producer men from 1921-1938 in a survey analysis of reported cancer deaths in England and Wales. No statistical significance testing was done and 30 occupational groups similarly had higher than predicted cancer deaths, including cabinet makers, drivers of horse-drawn vehicles, victuallers, and barmen.

Kawai et al. (1967) reported a statistically significant increase in lung cancer in 504 workers at a gas plant in a steel industry based on 6 cases. However, the expected number of lung cancer deaths in 1,451 person-years of observation was reported as 0.135 deaths and is clearly wrong. IARC (2012b) reports a nonsensical relative risk of 3,333 and then states: "Precision in the estimation of expected numbers was low." Clearly, this paper must be disregarded because it is erroneous.

Hansen et al. (1986) reported a statistically significant increase in lung cancer in 47 gas production workers employed for more than one year between 1911 and 1970 based on 7 cases in the gasworkers and 6 cases in the control group. This study is extremely limited in its scope.

Wu (1988) reported a statistically significant increase in lung cancer in 3,107 workers in 1971 at one of six gasification plants. However, IARC (2010b) states: “The short report does not allow an assessment of the validity of the study.”

The commenters note that little coal gasification has occurred in the United States since natural gas was discovered in the 1950's. The issue of cancer causation in coal gasification workers is a historical issue.

Coal tar pitch (paving and roofing): Baan et al. (2009) is cited by USEPA (2013) in Table 1-10, but this article presents no primary sources. The most recent IARC discussion of coal tar pitch in paving and roofing can be found in IARC (2012b) *Chemical Agents and Related Occupations* (Volume 100 F). The literature cited is listed below as the basis for the “sufficient” classification.

- Kennaway & Kennaway (1947)
- Kennaway & Kennaway (1951)
- Hammond, et al. (1976)
- Milham (1982)
- Pukkala (1995)
- Swaen & Slangen (1997)
- Stern et al. (2000)
- Schoenberg et al. (1987)
- Zahm et al. (1989)
- Morabia et al. (1992)
- Partanen & Boffetta (1994)
- Kauppinen et al. (2003)

Kennaway & Kennaway (1947) found an excess in the number of cancer deaths for pavers, street masons, concretors, and asphalters from 1921-1938 in a survey analysis of reported cancer deaths in England and Wales. No statistical significance testing was done and 30 occupational groups similarly had higher than predicted cancer deaths, including cabinet makers, drivers of horse-drawn vehicles, victuallers, and barmen.

Kennaway and Kennaway (1951) provides no information on roofers or pavers, so this is a mis-citation by IARC (2012).

Hammond, et al. (1976) found an increased number of lung cancer deaths compared to expected rates for pavers and roofers exposed for 20 or more years (22 deaths out of 4,215 workers) but not in workers exposed for 9-19 years (99 deaths out of 1724 workers). No statistical significance testing was done.

Milham (1982) reported that "Road Graders, Pavers, Machine Operators and Excavators" had elevated lung cancer risk but statistical significance was not presented or discussed.

Pukkala (1995) found a statistically significant increase in lung cancer risk in asphalt roofers in a national survey of 109,000 cases of cancer in Finland. Many other occupational groups had elevated lung cancer rates, including lawyers, wholesalers, door-to-door salesmen, and private sector managers.

Swaen & Slangen (1997) found no statistically significant increase in lung cancer risk in 866 roofers.

Stern et al. (2000) found a statistically significant increase in lung cancer mortality in a study of 11,144 members of the United Union of Roofers, Waterproofers, and Allied Workers. These workers were exposed to asphalt coal tar pitch, asbestos and fiberglass. Only 13% were exposed to coal tar pitch.

Schoenberg et al. (1987) found no statistically significant increase in lung cancer mortality in roofers and slaters in a case-control study of 763 lung cancer cases and 900 controls in New Jersey.

Zahm et al. (1989) found no statistically significant increase in lung cancer mortality in roofers or pavers/surfacers, material moving equipment operators in a case-control study of 4,431 lung cancer cases and 11,326 controls in Missouri.

Morabia et al. (1992) found no statistically significant increase in lung cancer mortality in roofers and slaters in a case-control study in 24 hospitals in nine metropolitan areas in the US.

Partanen & Boffetta (1994) was a meta-analysis of numerous case-control epidemiological studies. They reported a statistically significant increase in relative risk for lung cancer in roofers and "miscellaneous and unspecified bitumen/asphalt workers", but not in highway maintenance workers and pavers.

Bergdahl & Jarvholm (2003) found no statistically significant increases in lung cancer in 6,150 asphalt paving workers. No increases were seen except in the sub-group that started before 1954, and this increase was not statistically significantly increased.

Randem et al. (2003) reported an increase in lung cancer risk in 8,763 asphalt workers who worked for more than 3 months, but the risk was higher in those with 0-5 month's work history than in those with more than 6 months' work history. The authors concluded that smoking contributed to the observed increase in risk.

Stucker et al. (2003) reported that there was no statistically significant increase in lung cancer risk in 15,011 paving and road construction workers with greater than 6 months on the job.

Kauppinen et al. (2003) studied 2,642 bitumen workers, 2,552 building/ground construction workers and 382 other workers employed in road paving companies for at least six months before 1985. The cancer incidence study also includes temporary workers who worked only 3-6 months. The entire cohort of paving workers showed a small statistically significant increase in lung cancer mortality. When subgroups were analyzed, bitumen workers did not have a statistically significant increase in lung cancer mortality, but the building/ ground construction workers did. 4,815 (85%) of the cohort had no exposure to coal tar. Because the majority of the studied workers had no exposure to coal tar, this study should be excluded from any analysis of coal tar or coal tar pitch.

Coke production: Baan, et al. (2009) is cited by USEPA (2013) in Table 1-10, but this article presents no primary sources. The most recent IARC discussion of coal gasification is in IARC (2012) Chemical Agents and Related Occupations (Volume 100 F). The literature cited is listed below as the basis for the “sufficient” classification.

- Costantino et al., (1995)
- Wu (1988)
- Chau et al., (1993)
- Franco et al. (1993)
- Sakabe et al. (1975)
- Swaen et al. (1991)
- Buck & Reid (1956)
- Davies (1977)
- Hurley et al. (1983)
- Wu-Williams et al. (1993)

Costantino et al. (1995) reported a statistically significant increase in lung cancer deaths in 5,321 coke oven workers. The relative risk increased with increasing number of years as a coke-oven worker.

Wu, (1988) found a statistically significant increase in lung cancer deaths in 21,995 coke plant workers and also in coke oven workers. IARC (2012) reported that the description of methods was “insufficient.”

Chau et al. (1993) studied 535 coke oven plant workers and reported a statistically significant increase in lung cancer mortality in the total cohort and subgroups of workers who (a) were non-exposed and (b) who worked near the coke ovens. However, there was no statistically significant increase in lung cancer risk for

workers at the coke ovens, themselves. When smoking was considered, the non-exposed workers smokers and near coke oven smokers had statistically significant increases in lung cancer mortality, but the non-smokers did not.

Franco et al. (1993) found a barely significant increase (1.02-2.65) in lung cancer mortality in 538 coke oven workers from two plants who were employed from 1960-1985. Information on confounding exposures such as smoking and exposures from previous occupations was not available.

Sakabe et al. (1975) did not find any statistically significant increase in lung cancer mortality in 2,178 coke-oven workers in Japan who from 11 companies.

Swaen et al. (1991) did not find any statistically significant increase in lung cancer mortality in 5,659 workers who were employed for at least 6 months at any of three coke oven plants in 1945-1969.

Buck & Reid (1956) is mis-cited and is actually Reid and Buck (1956). In this study 5 cases of lung cancer death were expected and 4 were seen in coke oven workers who were so designated as their last job. When men were classified as coke oven workers if they worked at the coke ovens at any time during their employment, the expected number of lung cancer deaths was 10 and 14 were observed. Although no statistical tests were presented, the authors considered the results to be not significant: "The results obtained seem to imply that as far as recent experience in the coking industry goes, there is no great excess in cancer mortality in general nor in respiratory cancer in particular, even among the men who have worked on the ovens. This finding differs from the results of earlier work on the mortality of workers in the gas industry (Kennaway and Kennaway, 1947;: Doll, 1952) which suggested a twofold increase in respiratory cancer among men in that industry."

Davies (1977) did not find any statistically significant increase in lung cancer mortality in 601 coke oven workers who were active in 1954 at one of two plants.

Hurley et al. (1983) did not find any statistically significant increases in lung cancer mortality in 6,767 coke workers from 27 different plants who were active from 1966-1967 when the workers were categorized as "oven work," "part oven-work," or "non-oven work."

Wu-Williams et al. (1993) performed a case-control study of 965 lung cancer cases and 959 controls. They did not find any statistically significant increase in lung cancer incidence for women who stated on a questionnaire that they had exposures to coke oven emissions. This study is weak because exposure is self-assessed.

Diesel exhaust: Benbrahim-Tallaa et al. (2012) is cited by USEPA (2013) in Table 1-10, but this article presents no primary sources. The basis of the IARC's classification of diesel exhaust will be presented in Volume 105, which is not yet published.

Indoor emissions from household combustion of biomass fuel (primarily wood): Secretan et al. (2009) is cited by USEPA (2013) in Table 1-10, but this article has no discussion whatsoever about biomass

or wood. This is a mis-citation in USEPA (2013). IARC’s documentation for the *limited* classification for household combustion of biomass fuel (primarily wood) is in IARC (2010) *Household Use of Solid Fuels and High-temperature Frying* (Volume 95). Due to time limitations in preparing these comments to USEPA (2013), the cited papers on household combustion of wood could not be reviewed.

Indoor emissions from household combustion of coal: Secretan et al. (2009) is cited by USEPA (2013) in Table 1-10, but this article presents only one primary source: Straif et al. (2006). The IARC discussion of indoor combustion of coal is found in IARC (2012) *Personal Habits and Indoor Combustions* (Volume 100 E). Due to time limitations in preparing these comments to USEPA (2013), the cited papers on household combustion of coal could not be reviewed.

Soot (chimney sweeping): Baan, et al. (2009) is cited by USEPA (2013) in Table 1-10, but this article presents no primary sources. To find the basis for this classification, one needs to consult IARC (2012) Monograph 100F. IARC (2012) defined the mixture specifically as “soot, as found in occupational exposure of chimney sweeps.” The literature cited is listed below as the basis for the “sufficient” classification.

- Evanoff et al. (1993)
- Pukkala (1995)
- Pukkala et al. (2009)
- Kennaway & Kennaway (1947)
- Haldorsen et al. (2004)

Evanoff et al. (1993) found a statistically significant increase in relative risk of lung cancer mortality and incidence in 5,313 Swedish chimney sweeps that were working anytime from 1917 to 1980 when statistics were observed for workers who were exposed earlier in time and died on contracted cancer in the 1950s through ~1990. However, when only more recent statistics were evaluated, there was no statistically significant increase in cancer incidence or mortality compared to expected numbers. Adjustments for smoking were only made on the age class level. This study demonstrates that lung cancer in chimney sweeps, if associated with soot exposure, is a historical phenomenon.

Lung Cancer Findings of Evanoff et al. (1993)

Dates	Lung Cancer Mortality	Lung Cancer Incidence
1951-1990	Significant	
1983-1990	Not Significant	

1958-1987		Significant
1982-1987		Not Significant

Pukkala (1995) is mis-cited by IARC (2012). Specifically, the IARC document states:

“Evanoff et al. (1993) conducted large cohort study of Swedish chimney sweeps and found an excess of cancer of the lung, bladder, oesophagus and haematolymphatic organs; a study from Finland corroborated these findings (Pukkala, 1995).”

Pukkala (1995) does not corroborate the findings of Evanoff et al. (1993). Table 23 and appendix Table C-15 list the cancer incidence values for many different occupational groups, and there is no statistically significant increase in lung cancer in chimney sweeps.

Pukkala et al. (2009) performed a cancer registry study of 15 million people born between 1896 and 1960 in five Nordic countries. A statistically significant increase in lung cancer was seen in 5,498 chimney sweeps who worked from the turn of the century to 1960 in Denmark, Finland, Norway and Sweden based on 212 observed lung cancer cases. This is not an epidemiological study because no attempt was made determine how long a person worked in the self-identified occupational category.

Kennaway and Kenneway (1947) found no increase in lung cancer in 5,900 chimney sweeps from 1921-1938 in a survey analysis of reported cancer deaths in England and Wales.

Haldorsen et al. (2004) is reported in IARC (2012) to have performed a population cancer study with all men aged 25-64 in Norway and no statistically significant increase was seen in lung cancer in chimney sweeps.

Comparison of Information Reported in IARC Monographs to Results of the Cited Studies

Worker Group	IARC Reported Results for Lung Cancer	Actual Results for Lung Cancer
Aluminum Production		
Gibbs et al. (2007), Gibbs and Sevigny (2007a), Armstrong and Gibbs (2009)	Significant	Significant
Bjor et al. (2008)	Significant but no trend with PAH exposure levels	Significant but no trend with PAH exposure levels
Spinelli et al. (2006), Friesen et al. (2007)	Not significant, but significant trend with estimated BaP	Not significant; association with BaP exposure is not possible because only

**Comments on the
Toxicological Review of
Benzo[a]pyrene**

	exposure	particulate BaP was measured; only highest estimated BaP group statistically significant.
Friesen et al. (2009); Sim et al. (2009)	Not significant	Not significant
Mur et al. (1987)	Not significant	Not significant
Moulin et al. (2000)	Not significant	Not significant
Romundstad et al. (2000)	Not significant	Not significant
Rockett and Arena (1983)	Not significant	Not significant
Gibbs and Sevigny (2007b)	Not cited in text	Not significant for those first hired 1960-69 or 1970-79 Significant for those first hired from 1950-59
Giovanazzi & D'Andrea (1981)	Not cited in text	Not significant
Carta et al. (2004)	Not cited in text	Not significant
Carbon electrode manufacture		
Teta et al. (1987)	Not significant	Not significant
Moulin et al. (1989)	Not significant	Not significant
Gustavsson et al. (1995)	Not significant	Not significant
Liu et al. (1997)	Significant , but "...it is questionable how much of the excess risk may be attributed to exposures in carbon electrode manufacture."	Significant, but the group included an unknown number of aluminum smelter workers
Donato et al. (2000)	Not significant	Not significant

**Comments on the
Toxicological Review of
Benzo[a]pyrene**

Mori (2002)	Significant	Significant
Merlo et al. (2004)	Not significant	Not significant
Coal Gasification		
Doll et al. (1972)	Significant	Relative risk and significance level not provided
Berger & Manz (1992)	Significant	Significant
Martin et al. (2000)	Significant	Significant for highest exposure group but not significant for other exposure groups.
Kennaway & Kennaway (1947)	Significant	Significance level not provided
Kawai et al. (1967)	Significant, but "Precision in the estimation of expected numbers was low."	Paper must be disregarded. Expected number of lung cancer deaths in 1,451 person-years of observation was reported as 0.135 deaths.
Hansen et al. (1986)	Significant	Significant based on 7 cases in gasworkers compared to 6 in the controls.
Wu (1988)	Significant but IARC cannot evaluate the validity of the study.	Significant but the validity of the study cannot be validated.
Coal tar pitch (paving and roofing):		
Kennaway & Kennaway (1947)	Significance not discussed. Presented as increased risk.	Significance level not provided and category included workers besides pavers.
Kennaway & Kennaway (1951)	Significance not discussed. Presented as increased risk.	Mis-citation. No information on roofers, pavers, or any occupational groups.

**Comments on the
Toxicological Review of
Benzo[a]pyrene**

Hammond, et al. (1976)	Significance not discussed. Presented as increased risk.	Significance level not provided.
Milham (1982)	Significance not discussed. Presented as increased risk.	Significance not discussed in Milham (1982)
Pukkala (1995)	Significance not discussed. Presented as increased risk.	Significant
Swaen & Slangen (1997)	Significance not discussed. Presented as increased risk.	Not significant
Stern et al. (2000)	Significance not discussed. Presented as increased risk.	Significant
Schoenberg et al. (1987)	Not significant	Not significant
Zahm et al. (1989)	Not significant	Not significant
Morabia et al. (1992)	Not significant	Not significant
Partanen & Boffetta (1994)	Significant meta-analysis for roofers	Significant meta-analysis for roofers, but not pavers
Bergdahl & Jarvholm (2003)	Not significant	Not significant
Randem et al. (2003)	Not significant	Significant but risk was greater in those with less time working.
Stucker et al. (2003)	Not significant	Not significant
Kauppinen et al. (2003)	Significant	Study should be excluded because 85% of the cohort had no exposure to coal tar pitch.
Coke production		
Costantino et al. (1995)	Significant	Significant
Wu (1988)	Significant	Significant, but IARC (2012) states that the methods were

		“insufficient.”
Chau et al. (1993)	Significant	Significant for “near coke oven” workers but not for “coke oven” workers
Franco et al. (1993)	Significant	
Sakabe et al. (1975)	Not significant	Not significant
Swaen et al. (1991)	Not significant	Not significant
Buck & Reid (1956)	Significant	Not significant
Davies (1977)	Not significant	Not significant
Hurley et al. (1983)	Not significant	Not significant
Wu-Williams et al. (1993)	Not reported	Not significant
Soot (chimney sweeping)		
Evanoff et al. (1993)	Significant	Significant for 1950’s -1980’s Not significant for 1980’s
Pukkala (1995)	Significant	Not significant
Pukkala et al. (2009)	Significant	Significant
Kennaway & Kennaway (1947)	No increase	No increase
Haldorsen et al. (2004)	Not significant	Not significant

USEPA (2013) cites IARC documents that summarized the human evidence of increased lung cancer rates in various industry worker populations. Despite the fact that BaP is not listed as a causal agent in these studies, the studies themselves provide marginal evidence that lung cancer was increased in these workers who were exposed to complex mixtures of chemicals. In fact, there are twice as many studies cited showing no increased risk of lung cancer than there are positive studies. Specifically, of 38 cited studies that were reported to be adequate in quality, 25 were negative studies and 13 were positive studies.

Summary and Evaluation

2. Does USEPA's hazard assessment of noncancer human health effects of benzo[a]pyrene clearly integrate the available scientific evidence (i.e., human, experimental animal, and mechanistic evidence) to support the conclusion that benzo[a]pyrene poses a potential hazard to the developing fetus; the nervous system in the developing fetus; the male and female reproductive systems; and the immune system?

COMMENT: No comment due to time constraints. No comment does not mean tacit agreement.

3. Does USEPA's hazard assessment of the carcinogenicity of benzo[a]pyrene clearly integrate the available scientific evidence to support the conclusion that under USEPA's Guidelines for Carcinogen Risk Assessment (U.S. USEPA, 2005), benzo[a]pyrene is "carcinogenic to humans" by all routes of exposure?

COMMENT:

As noted in comment C-1, USEPA (2013) has mischaracterized the weight of evidence by indicating that BaP causes skin cancer in humans.

4. Does USEPA's hazard assessment of the mode of action for carcinogenicity of benzo[a]pyrene clearly integrate the available scientific evidence to support the conclusion that a mutagenic mode of action is the primary mode of action of benzo[a]pyrene-induced carcinogenicity?

COMMENT: No comment due to time constraints. No comment does not mean tacit agreement.

D. Dose-Response Analysis

Oral Reference Dose (RfD)

Several hazards were identified for oral exposure to benzo[a]pyrene. Studies and effects within each hazard (i.e., developmental, reproductive, and immunotoxicity) were evaluated and the most relevant, informative studies and effects were selected for dose-response analysis, where data were amenable, for consideration in deriving an RfD.

1. Please comment on whether the evaluation and selection of studies and effects for the derivation of candidate values to consider for the RfD are scientifically supported and clearly described. Specifically,

please comment on the selection of the following studies and effects for dose-response analysis. Please identify and provide the rationale for any other studies or effects that should be considered.

- a. Developmental toxicity: Chen et al. (2012) [neurodevelopmental changes]; Jules et al. (2012) [cardiovascular effects]

COMMENT:

USEPA (2013) relies exclusively on the study of Chen et al. (2012) for the RfD derivation, but this study is flawed in many ways and should be excluded from the RfD derivation process. Both Chen et al. (2012) and Jules et al. (2012) are evaluated below and it is recommended that the developmental toxicity endpoint be represented by the results of Jules et al. (2012).

Chen et al. (2012): Chen et al. (2012) performed 72 separate comparison tests in their study of Sprague-Dawley rats treated with BaP on postnatal day (PND) 5 through PND 11 by gavage in peanut oil. USEPA (2013) has focused on just three of these tests that reportedly showed statistically significant results. With $p=0.05$ as the significance criterion, one would expect 4 statistically significant differences due solely to chance when so many separate observations are made. The following table lists all of the tested performed by the investigators.

Tests Performed By Chen et al. (2012)

Test Performed	Subtest Performed	Number of Observation Groups
Developmental Milestones		
Body weight		PND 5 PND 6 PND 7 PND 8 PND 9 PND 10 PND 11 PND 36 PND 71
Incisor eruption		Day observed
Eye opening		Day observed
Fur development		Day observed
Testis decent		Day observed
Vaginal opening		Day observed
Neonatal Sensory and Motor Development Tests		
Surface righting reflex test (track 1 animals)		PND 12 PND 14 PND 16 PND 18
Negative geotaxis test		PND 12

**Comments on the
Toxicological Review of
Benzo[a]pyrene**

(track 2 animals)		PND 14 PND 16 PND 18
Cliff aversion test (track 1 animals)		PND 12
Forelimb grip strength test (track 2 animals)		PND 12
Open-field test	Horizontal Movement	PND 18 (track 1 animals) PND 20 (track 2 animals) PND 34 (track 3 animals) PND 69 (track 4 animals)
Open-field test	Vertical Movement, Rearing	PND 18 (track 1 animals) PND 20 (track 2 animals) PND 34 (track 3 animals) PND 69 (track 4 animals)
Elevated plus maze	Latency Time of the first entry into an open arm	Male PND 35 (track 3 animals) Male PND 70 (track 4 animals) Female PND 35 (track 3 animals) Female PND 70 (track 4 animals)
Elevated plus maze	Time Spent in the Open Arm	Male PND 35 (track 3 animals) Male PND 70 (track 4 animals) Female PND 35 (track 3 animals) Female PND 70 (track 4 animals)
Elevated plus maze	Number of Entries into the Open Arms	Male PND 35 (track 3 animals) Male PND 70 (track 4 animals) Female PND 35 (track 3 animals) Female PND 70 (track 4 animals)
Elevated plus maze	Number of Entries into the Closed Arms	Male PND 35 (track 3 animals) Male PND 70 (track 4 animals) Female PND 35 (track 3 animals) Female PND 70 (track 4 animals)
Morris water maze	Escape Latency	Adolescent males PNP 36 (track 3 animals) PNP 37 (track 3 animals) PNP 38 (track 3 animals) PNP 39 (track 3 animals) Adolescent females PNP 36 (track 3 animals) PNP 37 (track 3 animals) PNP 38 (track 3 animals) PNP 39 (track 3 animals) Adult males PNP 71 (track 4 animals) PNP 72 (track 4 animals) PNP 73 (track 4 animals) PNP 74 (track 4 animals) Adult females PNP 71 (track 4 animals) PNP 72 (track 4 animals) PNP 73 (track 4 animals) PNP 74 (track 4 animals)
Morris water maze	Number of Times Animal	Male PNP 40 (track 3 animals)

	Crossed Original Platform in Probe Test	Male PNP 75 (track 4 animals) Female PNP 40 (track 3 animals) Female PNP 75 (track 4 animals)
Morris water maze	Time Spent in the Target Quadrant	Male PNP 40 (track 3 animals) Male PNP 75 (track 4 animals) Female PNP 40 (track 3 animals) Female PNP 75 (track 4 animals)

The rats used in this experiment were used for multiple tests that occurred over seven day periods, so it is unclear whether the reported performance on the test was due to BaP exposure or due to the stress and strain of the experimental regimen.

Neurobehavioral Testing in Chen et al. (2012)

Post Natal Day	Track 3 Rats 20/group x 4 groups	Track 4 Rats 20/group x 4 groups
Day 34	Open-Field Test	
Day 35	Elevated Plus Maze	
Day 36	Morris Water Maze	
Day 37	Morris Water Maze	
Day 38	Morris Water Maze	
Day 39	Morris Water Maze	
Day 40	Morris Water Maze	
Day 69		Open-Field Test
Day 70		Elevated Plus Maze
Day 71		Morris Water Maze
Day 72		Morris Water Maze
Day 73		Morris Water Maze
Day 74		Morris Water Maze
Day 75		Morris Water Maze

Chen et al. (2012) tested male and female rats on eight different occasions in the Morris Water maze test to see what their escape latencies were. There was no effect of BaP treatment at any dose on PND 36, 37, 38, 71, 72, or 73, but there was a reportedly significant effect on days 39 and 74 for both males and females. USEPA (2013) totally ignores the fact that in more times than not, no significant differences were seen. They focus, instead, on the minority of times when the tests showed significant differences at the $p < 0.05$ level.

When performing benchmark dose modeling, USEPA (2013) merged the data for the males and females despite the fact that the raw data are not presented in the paper to allow group statistics to be calculated. It is not reported whether USEPA used PND 39 data or PND 74 data or whether they somehow combined the data from adolescent and adult rats.

USEPA (2013) also focused on the data for the “time spent in the target quadrant” test, again somehow merging the male and female data without explanation. Data are presented in the paper separately for male PND 40, male PND 75, female PND 40 and female PND 75. It is not reported what specific data were modeled and whether data from day 40, day 75 or both days were merged. Again, the information reported in the paper does not support the merging of data.

USEPA (2013) focused in on the results of the elevated maze tests and specifically the number of open entries by females at PND 70. The benchmark dose low ($BMDL_{10}$) from this single test is the basis of the USEPA’s proposed RfD. This maze test is a test of anxiety in rodents. If the rodents are anxious or fearful, they will avoid the open arms and choose the closed arms compared to control animals.

No statistically significant differences were seen in the number of entries into either the open or closed arms of the maze. On PND 35, there were no differences between treatment groups with regard to their entry into either the open arms or the closed arms. However, only on PND 70, the dosed animals (high dose for males, middle and high dose for females) entered the open arms more frequently and entered the closed arms less frequently. USEPA (2013) reports that this is “decreased anxiety-like behavior.” Given that this test is used in pharmacology to test the efficacy of anxiety reducing drugs, it is unclear why *decrease in anxiety* is labeled an “adverse effect.”

More importantly, this test is well documented to be a subjective test the results of which are highly influenced by (1) housing of animals, (2) the scoring method used, (3) the construction of the maze, and (4) pre-test manipulation (Hogg, 1996).

Specifically, Hogg (1996) states that single housing of the animals from 30 minutes to seven days is routinely done by investigators using this test. Chen says that: “Rats were placed in the quiet experimental room for a number of 30 min prior to testing.” They do not state whether the animals were in single or multiple cages. They also do not state how long the animals were in the quiet room before testing. Given that the testing took four hours to complete on any given day, it seems likely that all of the animals were brought into the quiet room at the same time and 30 minutes later the testing was started. If this is the case, the first group of animals had a 30 minute rest before the test and the other groups had progressively longer periods of quiet rest up to several hours. Clearly, differing periods of resting would affect the rats’ anxiety

states, which is the subject of the test, and the Chen et al. (2012) is deficient by not discussing this issue.

The scoring method used seems to be appropriate with an “entry” being counted if all four paws were in the arm.

The construction of the maze also affects the results. Hogg (1996) states that the addition of ledges or raised lips around the edges of the open arms can affect study results. Chen et al. (2012) is silent on this issue, but because of its importance, it should have been discussed.

Pre-test manipulation is critical. According to Hogg (1996), “repeated handling of animals for several days before experimentation serves to habituate them to the stresses to which they are commonly subjected immediately before plus-maze testing...” and “It would be naïve to assume that animals will respond in the same way on exposure to the [elevated plus-maze] EPM irrespective of their manipulations beforehand.”

No information is provided in Chen et al. (2012) about the above factors that affect the anxiety state of the rats prior to and during the elevated plus maze test. It is known that the results are highly variable because the control animals varied in their responses over the track 3 and track 4 animals:

	<u>Number of Open Arm Entries</u>
Track 3, adolescent males (PND35)	~12
Track 4, adult males (PND70)	~14
Track 3, adolescent females (PND35)	~11
Track 4, adult females (PND70)	~10

Given the small number of open arm entries in a 5 minute period, this variability in the control baseline is troubling because it affects the statistical significance testing in an unknown manner.

In addition, the *number of open arm entries* is a poor metric for increased or decreased anxiety state. According to Hogg (1996), “expression of the open arm data as percentages of the total number of arm entries (to give % number of open arm entries; %no) or total time spent (to give % time on open arms; %t) on either the open or closed arms corrects for overall changes in exploration of the maze and helps to reduce activity-induced artifacts.” Chen et al. (2012) did not normalize the data as recommended by Hogg (1996) and instead reported the raw data as their metric.

“Decreased anxiety-like behavior” should not be considered an appropriate endpoint for RfD derivation because many details are not provided in this study about the manner in which animals were handled which can affect anxiety levels and because this is not an adverse effect.

A developmental toxicity study discussed by USEPA (2013) but not used for dose-response assessment is Jules et al. (2012). The Jules study is superior to Chen et al. (2012) and should be used for RfD derivation.

Jules et al. (2012): Jules et al. (2012) dosed pregnant Long Evans Hooded rats with BaP at 0.15, 0.30, 0.60, and 1.2 mg/kg-day by gavage on embryonic days 14 through 17. According to the authors: "There were no significant effects of in utero exposure to B(a)P on the number of pups born per litter or in pre-weaning growth curves." There were no effects on the number of pups per litter or on the body weights of offspring on post natal days 0, 3, 5, 7, 9, 11, 13, or 15. Heart rate, systolic blood pressure and diastolic blood pressure were measured in the 4-5 animals from the control group and the two highest dose groups but not the two lower dose groups. Heart rate is higher in the 0.6 mg/kg-day group and lower in the 1.2 mg/kg-day group, so this endpoint cannot be used to define a clear NOAEL or LOAEL. Systolic blood pressure and diastolic blood pressure both increase in a dose-responsive manner in the two dose groups, and USEPA (2013) defined the 0.6 mg/kg-day dose group as a LOAEL for increase in systolic and diastolic blood pressure. Note that the NOAEL was likely 0.30 mg/kg-day but this dose group was not studied for changes in blood pressure. Benchmark dose modeling was not done for either.

This study should be rejected for dose-response assessment because key information is missing and erroneous data are reported. The study does not define how many animals were tested for blood pressure readings. Different facts are reported in different places in the paper, for example:

- (a) "...using littermates from four to five different litters within an experimental group would be sufficient to detect a significant difference."
- (b) "Since the litter is the statistical unit and the sampling was from at least 4 to 5 individual litters with in an experimental group..."
- (c) "To assess whether *in utero* B(a)P exposure affected blood pressure in offspring rats, the systolic blood pressure of conscious-nonanesthetized animals (n=5-6/litter/group) was measured on post natal day 53."
- (d) "Values are expressed as the mean +/- [standard error of the mean] SEM of three separate experiments in which blood pressure measurements were averaged from four-five rat offspring."

A paper with so many discrepancies in the reporting of critical data should not be used for dose-response assessment.

There are also significant errors in the reporting. On Figure 3, the SEM is shown as a bar above the mean of the measurements. Clearly, the SEM increases as one moves from control to low dose to higher dose. In Table 3B, however, the SEMs are listed as 1.2, 45, and 2.4 in the same order. Clearly, 45 is an error. In addition, in the text the blood pressure for the control group is listed as 131.5 mmHg with an SEM of 5.8 mmHg. This does not agree with the data in Table 3B which is given as 131.6 +/-1.2. Is the SEM 1.2 or 5.8? If the data in table 3B were not SEM, but standard deviation instead, then one can calculate the *n* for the

control group as 23. If they tested 5 per litter times 4 litters, *n* would be 20. If they tested 6 per litter times 4 litters, *n* would be 24. If they tested 5 per litter times 5 litters, *n* would be 25.

In conclusion, one cannot determine from this paper if the error values presented are SEM or standard deviation (SD) and there is no way to determine how many animals were tested from each litter, how many litters were tested, and how many total animals were tested. This information is required to perform benchmark dose modeling, so this is apparently the reason that USEPA (2013) is silent on benchmark dose modeling and uses a NOAEL/LOAEL approach instead.

Instead, this paper is perhaps not suitable for use in dose-response modeling, because there is no way to independently verify the statistical test results reported by the authors. If, on the other hand, USEPA could obtain the raw data from the authors and then independently perform statistical significance testing and subsequent benchmark dose modeling, this paper may have greater utility in dose-response assessment of BaP.

- b. Reproductive toxicity: Xu et al. (2010) [decreased ovary weights]; Zheng et al. (2010) [decreased intratesticular testosterone]; Mohamed et al. (2010) [decreased sperm count and motility]; and Gao et al. (2011) [increased cervical hyperplasia].

COMMENT:

USEPA (2013) quantitatively assesses studies by Xu et al. (2010), Zheng et al. (2010), Mohamed et al. (2010), and Gao et al. (2011) for the reproductive endpoint and then rejects Zheng et al. (2010), Mohamed et al. (2010), and Gao et al. (2011) to focus the RfD entirely on one endpoint from the study of Xu et al. (2010). Each of these four studies is evaluated below, and it is recommended that point of departure doses and candidate RfDs be derived from each and then averaged to represent the reproductive endpoint. USEPA (2013) should not focus entirely on the one candidate RfD simply because it is the worst case.

Xu et al. (2010): Xu et al. (2010) observed decreased ovary weights in female Sprague-Dawley rats given BaP 5 mg/kg and 10 mg/kg in corn oil intragastically every other day for 60 days. Statistically significant decreases in absolute ovary weights were seen at both BaP doses but the ovary/body weight ratio was only decreased in the highest dose group.

With regard to other endpoints, 5 mg/kg-day dose is a NOAEL for reproductive effects as shown in Figure 1 of the paper, which shows that the control animals and the 5 mg/kg-day animals were identical in their estrous cycles. Only at 10 mg/kg-day were significant difference seen between treated animals and controls.

The same is seen in Figure 3, which shows the effects of BaP on the ovarian follicle populations and corpora lutea in the treated animals after 90 days. There were no differences between control animals and animals treated with 5 mg/kg-day BaP. Only 10 mg/kg-day caused any significantly different effects.

USEPA's benchmark dose modeling of the Xu et al. (2010) ovary weight data resulted in a BMDL₁₀ of 1.49 mg/kg-day after adjusting for every-other-day dosing. However, 5 mg/kg-day is a NOAEL for other measures of reproductive toxicity, and these NOAELs should also be considered in the dose-response modeling. Relevant results from Xu et al. (2010) are reported in the table below.

Results from Xu et al. (2010)

Reproductive Endpoint	NOAEL	LOAEL	USEPA (2013) Focus
Decreased Ovary Weight	None	5 mg/kg-day	BMDL ₁₀ defined as 1.49 mg/kg-day.
Estrous cycles	5 mg/kg-day	10 mg/kg-day	Not discussed or modeled
Ovarian follicle populations	5 mg/kg-day	10 mg/kg-day	Not discussed or modeled
Corpora lutea	5 mg/kg-day	10 mg/kg-day	Not discussed or modeled

Zheng et al. (2010): Zheng et al. (2010) gave male Sprague-Dawley rats 1 or 5 mg/kg-day of BaP every day for 90 days in corn oil via gavage. Intratesticular testosterone levels were not significantly decreased in BaP-dosed animals at either dose level at day 30. At day 90, the testosterone levels were statistically significantly decreased only at the 5 mg/kg-day dose level. Thus, 1 mg/kg-day was a NOAEL for male reproductive effects in this study. The results of Zheng et al. (2010) should be included in the RfD calculation. Relevant results from this study are presented in the following table.

Results from Zheng et al. (2010)

Reproductive Endpoint	NOAEL	LOAEL	USEPA (2013) Focus
Intratesticular testosterone levels at day 30	5 mg/kg-day	None defined	Not discussed or modeled
Intratesticular testosterone levels at day 90	1 mg/kg-day	5 mg/kg-day	NOAEL defined as POD

Mohamed et al. (2010) administered BaP at 1 or 10 mg/kg-day in corn oil by gavage daily for six weeks. F1, F2, and F3 offspring were treated in a similar manner and tested for male fertility parameters. There were not statistically significant differences between treated and control animals over the generations for testicular morphology. A decrease in the percentage of seminiferous tubules containing elongated spermatids was significantly different from the controls in the 10 mg/kg-day group. Sperm count was significantly lower in the 1 and 10 mg/kg-day groups in the F0 and F1 generations. No effects were seen on sperm count in the F3 generation. Sperm motility was significantly lower in the 10 mg/kg-day groups in the F0 generation. There were no effects on the sperm motility in the F2 and F3 generations. The LOAEL of 1 mg/kg-day for effects on sperm count should be considered in the derivation of the RfD as noted below in a comment on RfD derivation.

- c. Immunotoxicity: Kroese et al. (2001) [decreased thymus weights]; DeJong et al. (1999) [decreased serum IgM and IgA levels and number of B cells]

COMMENT:

USEPA (2013) quantitatively assesses studies by Kroese et al. (2001) and DeJong et al. (1999) for the immunotoxicity endpoint and then rejects Kroese et al. (2001) to focus the RfD entirely on one endpoint from the study of DeJong et al. (1999). Each of these two studies is evaluated below, and it is recommended that point of departure doses and candidate RfDs be derived from each and then averaged to represent the immunotoxicity endpoint. USEPA (2013) should not focus entirely on the one candidate RfD simply because it results in a worst case RfD. USEPA (2013) should quantitatively consider the entire scientific weight of evidence in deriving RfDs.

Kroese et al. (2001): Kroese et al. et al. (2001) treated 10 male and 10 female Wistar rats for five days per week for three months with BaP at doses of 0, 3, 10, or 30 mg/kg-day by gavage in soy oil. In both males and females, the highest dose (30 mg/kg-day) is a LOAEL, but USEPA (2013) modeled only the female data and not the male data. Relevant results from Kroese et al. et al. (2001) are presented in the following table.

Results from Kroese et al. (2001)

Immunological Endpoint	NOAEL	LOAEL	USEPA (2013) Focus
Thymus weight, males	10 mg/kg-day	30 mg/kg-day	Discussed but not modeled
Thymus weight, females	10 mg/kg-day	30 mg/kg-day	BMDL ₁₀ defined as 7.6 mg/kg-day

Benchmark dose modeling was performed by the commenters on the Kroese et al. (2001) data, and the best fitting model for female thymus weight gives a BMDL₁₀ of 7.65 mg/kg-day. Male thymus weight data were not amenable to benchmark dose modeling as noted in the following table. It is recommended that the

candidate RfD from Kroese et al. (2001) be used quantitatively in RfD derivation as noted in comments below.

Benchmark Dose Modeling of Results from Kroese et al. (2001)

Immunological Endpoint	USEPA (2013)	These Comments
Thymus weight, males	Not amenable to BMDM	Not amenable to BMDM
Thymus weight, females	BMDL ₁₀ = 7.6 mg/kg-day (linear model) (USEPA rounded down instead of up)	BMDL ₁₀ = 7.65 mg/kg-day (linear model)

De Jong et al. (1999): The study by De Jong and colleagues was solely used to define the candidate RfD for the immunotoxicity endpoint, and USEPA (2013) focused entirely on one endpoint among many under study. This study is evaluated below and appropriate endpoint and quantitative measures of effect are recommended.

De Jong et al. (1999) dosed male Riv:Tox Wistar rats for 35 days with BaP (3,10, 30, or 90 mg/kg-day) by gavage in soybean oil for 5 days a week. They made 43 comparisons among the groups, including body weight, weights of brain, heart, kidney, liver, thymus, spleen, mandibular lymph nodes, mesenteric lymph nodes, popliteal lymph nodes, thymus cortex surface area, cortex to thymus weight ratio, medulla to thymus weight ratio, spleen cells (total, B, T, Th, and Ts), bone marrow cells, spleen cell distribution (B, T, Th, and Ts), white blood cells, red blood cells, hemoglobin, Ht, MCV, MCH, MCHC, RDW, PLT, Q-index, lymphocytes, neutrophils, eosinophils, basophils, monocytes, IgM, IgG, IgA, and IgE . For some of these parameters, there were not significant changes seen. Focusing on the immunological endpoints, there are an almost equal number of comparisons that we not statistically significantly different from controls as there were statistically significant comparisons. Relevant results from De Jong et al. (1999) are shown in the following table.

Immunological Effects in De Jong et al. (1999)

Not Significant	Significant
Spleen weight	Thymus weight
Spleen B cells	Mandibular lymph nodes weight
Spleen T cells	Mesenteric lymph node weight

Spleen Th cells	Popliteal lymph node weight
Spleen Ts cells	Thymus cortex surface area
Spleen T cell distribution	Cortex to thymus weight ratio
Spleen Th cell distribution	Medulla to thymus weight ratio
Spleen Ts cell distribution	Total spleen cells
Circulating neutrophils	Spleen B cells*
Circulating basophils	Bone marrow cells
Circulating monocytes (no dose-response)	Spleen B cell distribution
Serum IgM (no dose-response)*	Circulating white blood cells
Serum IgG	Circulating lymphocytes
Serum IgE	Circulating eosinophils
	Serum IgA*

*Chosen by USEPA (2013) for candidate RfD derivation

The summary of a detailed evaluation of De Jong et al. (1999) is shown in the following table.

Selected USEPA Immunological Rodent Points of Departure from De Jong et al. (1999)

Immunological Endpoint	NOAEL	LOAEL	USEPA (2013) Focus
Decreased serum IgM	No clear NOAEL because effects exhibited no dose-response	No clear LOAEL because effects exhibited no dose-response	NOAEL of 10 mg/kg-day (Adjusted NOAEL of 7.1 mg/kg-day) defined
Decreased IgA	30 mg/kg-day	90 mg/kg-day	NOAEL of 30 mg/kg-day (Adjusted NOAEL of 21.4

			mg/kg-day) defined
Decreased spleen B cells	30 mg/kg-day	90 mg/kg-day	NOAEL of 30 mg/kg-day (Adjusted NOAEL of 21.4 mg/kg-day) defined

USEPA (2013) should exclude the IgM response from the candidate RfD derivation process, because the IgM levels are statistically significantly depressed in the 30 mg/kg-day group, but the IgM levels are not statistically significantly different from the control group in the 90 mg/kg-day group. There is no clear dose-response for this effect, so a NOAEL or a LOAEL cannot be derived for this endpoint.

2. Benchmark dose (BMD) modeling was applied to derive points of departure (POD) for the candidate values when possible. Has the BMD modeling been appropriately conducted? Are the benchmark responses (BMR) selected for use in deriving the PODs scientifically supported and clearly described? When BMD modeling was not possible a NOAEL/LOAEL approach was used to calculate candidate values. Please comment on whether these approaches are scientifically supported and clearly described.

COMMENT:

The commenters agree that the benchmark dose modeling is a reasonable approach when possible. According to USEPA (2013), the data of Jules et al. (2012), Zheng et al. (2010), Mohamed et al. (2010), and De Jong et al. (1999) were not amenable to benchmark dose modeling. If true, then the NOAEL/LOAEL approach is acceptable.

The commenters disagree with USEPA's selection of endpoints from a large number of studies and a large number of comparisons within some of the studies. For instance, the Chen et al. (2010) study performed 77 different comparisons, and USEPA (2013) focused on three and then dismissed two of those to put forth one comparison from this complex study for candidate RfD derivation.

3. Please comment on the rationale for the selection of the uncertainty factors (UFs) applied to the PODs for the derivation of the candidate values. Are the UFs appropriate, based on the recommendations described in Section 4.4.5 of A Review of the Reference Dose and Reference Concentration Processes (USEPA, 2002), and clearly described? If changes to the selected UFs are proposed, please identify and provide scientific support for the proposed changes.

COMMENT:

The commenters agree with the usual Uncertainty Factors, but disagree that all candidate RfD require a "Database Uncertainty Factor" of 3. BaP has one of the most complete sets of toxicological data of any

chemical ever studied by toxicologists. There is no need to account for “database uncertainty,” and this factor of 3 should be removed from all RfD derivations.

- From the candidate values, an organ/system-specific reference value was selected for each hazard (developmental, reproductive, and immunotoxicity). USEPA concluded that these values best represented the hazards based on considerations of weight of evidence, uncertainty, and sensitivity. Please comment on whether the selection of the organ/system-specific reference value is scientifically supported, appropriate for development of a chronic RfD, and is clearly described. Please identify and provide the rationale for any other values that should be considered.

COMMENT:

The commenters disagree with the manner in which studies with multiple toxicological comparisons were mined to find test/control comparisons that gave the lowest BMDL₁₀ or NOAEL. At every step, USEPA (2013) chose the worst case rather than consider the overall weight of evidence from within or among available studies. For instance, with the data of Xu et al. (2010) an animal point of departure dose of 1.5 is chosen for decreased ovary weight, but the higher point of departure doses that result from a focus on estrous cycles or ovarian follicle populations or corpora lutea were ignored. Similarly, with De Jong et al. (1999), USEPA (2013) ignored a dozen comparisons that showed little or no effects of BaP dosing, but focused on the IgM response, which was not appropriate because of a lack of dose-response. Even when USEPA (2013) presented three endpoints within that single study for consideration, they ignored the two that gave higher points of departure doses. The same worst case focus was used when comparing studies. For instance, USEPA (2013) ignored a point of departure dose from Kroese et al. (2001) and chose a point of departure dose from De Jong et al. (1999) because the dose from Kroese et al. (2001) was higher.

Instead of ignoring all of the candidate RfDs except for the one with the lowest point of departure dose, USEPA (2013) should, instead, consider and *use* more of the results from the available studies to compute candidate RfDs and then propose as the final RfD an *average value from multiple studies*. This is the more scientifically neutral approach that takes into account the entire scientific weight of evidence.

In terms of Point of Departure doses, the following table summarizes the above comments and proposes for USEPA’s consideration a revised table of Point of Departure Doses for RfDs for three endpoint categories.

Appropriate Point of Departure Doses for Candidate Reference Doses

Endpoint Category	Endpoint Subcategory	Comments	Reference	NOAEL, LOAEL or BMDL ₁₀	POD _{HED}
Developmental	Neurobehavioral Number of Open Arm	One test in one sex on one test date out of 77 tests	Chen et al. (2011)	BMDL₁₀ 0.09	0.09*

Comments on the
Toxicological Review of
Benzo[a]pyrene

	Entries	many of which showed no effects. Effect noted was a measure of <i>reduced</i> anxiety, which should not have been labeled <i>adverse</i> .			
	Cardiovascular		Jules et al. (2012)	LOAEL = 0.6	0.15
Reproductive	Decreased ovary weight		Xu et al. (2010)	BMDL ₁₀ = 1.5	0.37 ¹
	Estrous cycles/ovarian follicle populations.corpora lutea	Ignored by USEPA (2013)	Xu et al. (2010)	NOAEL = 2.5	0.61 ¹
	Decreased intratesticular testosterone at day 90		Zheng et al. (2010)	NOAEL = 1	0.24
	Decreased sperm count & motility		Mohamed et al. (2010)	LOAEL = 1	0.24
	Cervical epithelial hyperplasia		Gao et al. (2011)	BMDL ₁₀ =0.37	0.06
Immunological	Decreased thymus weight, females		Kroese et al. (2001)	BMDL ₁₀ =7.65	1.87
	Decreased serum IgM	Because of no dose-response, this endpoint is inappropriate	De Jong et al. (1999)	NOAEL= 7.1 *	4.73*

	Decreased serum IgA & Decreased # of B cells		De Jong et al. (1999)	NOAEL = 21.4	5.22
--	--	--	-----------------------	--------------	------

* Should be excluded from candidate RfD derivation

¹ These two PODs should be averaged to give one POD for Xu et al. (2010).

The summarization of the above table appears below to show the human equivalent PODs for each appropriate study within each endpoint category.

Human Equivalent Point of Departure Doses for Benzo(a)pyrene

Endpoint Category	Reference	POD _{HED}
Developmental	Jules et al. (2012)	0.15
Reproductive	Xu et al. (2010)	0.49*
	Zheng et al. (2010)	0.24
	Mohamed et al. (2010)	0.24
	Gao et al. (2011)	0.06
Immunological	Kroese et al. (2001)	1.87
	De Jong et al. (1999)	5.22

* Average of two PODs from this study.

- The proposed overall RfD was based on neurodevelopmental changes observed by Chen et al. (2012). This value was selected based on the confidence in and sensitivity of the reference value. Please comment on whether the selection of this RfD is scientifically supported and clearly described. Please identify and provide the rationale for any other values that should be considered.

COMMENT:

Please see comment above for a discussion of the Chen et al. (2012) study. This study defines decreased anxiety in rats as an adverse effect based on the results of testing at one time point in one sex. Rodents tested at other time points were not considered. Chen et al. (2012) should be rejected for RfD definition.

Instead, USEPA (2013) should use the weight of evidence of the entire data base and propose a RfD that is an average value from the available candidate RfDs. The commenters propose for USEPA's consideration that the average of each appropriate Point of Departure dose within each endpoint category should be calculated and used as the POD for the category as noted in the following table.

Candidate Reference Doses for Benzo(a)pyrene

Endpoint/Reference	POD _{HED}	Type	UF _A	UF _H	UF _I	UF _S	UF _D	Composite UF	RfD
Developmental									
Cardiovascular Jules et al. (2012)	0.15	LOAEL	3	10	10	1	1	300	5x10 ⁻⁴
Reproductive									
Ovary weight, Estrous cycles/ovarian follicle populations.corpora lutea Xu et al. (2010)	0.49	BMDL ₁₀ and NOAEL	3	10	1	10	1	300	1.6x10 ⁻³
Intratesticular testosterone Zheng et al. (2010)	0.24	NOAEL	3	10	1	10	1	300	8x10 ⁻⁴
Sperm count/motility Mohamed et al. (2010)	0.24	LOAEL	3	10	10	10	1	3000	8x10 ⁻⁵
Cervical hyperplasia Gao et al. (2011)	0.06	BMDL ₁₀	3	10	1	10	1	300	2x10 ⁻⁴
Immunological									
Thymus weight Kroese et al. (2001)	1.87	BMDL ₁₀	3	10	1	10	1	300	6.2x10 ⁻³

IgA & Spleen B cells	5.22	NOAEL	3	10	1	10	1	300	7.3×10^{-4}
De Jong et al. (1999)									

The commenters propose that reasonable RfDs can be derived for each of the three endpoint categories defined by USEPA and that the final RfD should be the average of the endpoint RfDs, which is 1.6×10^{-3} mg/kg-day.

Endpoint Category Reference Doses for Benzo(a)pyrene and Proposed Average Reference Dose

Endpoint	Category RfD	Average RfD
Developmental	5×10^{-4}	
Reproductive	6.8×10^{-4}	
Immunological	3.5×10^{-3}	
Average		1.6×10^{-3}

Inhalation Reference Concentration (RfC)

Several hazards were identified for inhalation exposure to benzo[a]pyrene. Studies and effects within each hazard (i.e., developmental and reproductive) were evaluated and the most relevant, informative studies were selected for dose-response analysis, where data were amenable, for consideration in deriving an RfC.

6. Please comment on whether the evaluation and selection of studies and effects for the derivation of candidate values to consider for the RfC are scientifically supported and clearly described. Specifically, please comment on the selection of the following studies and effects for dose-response analysis. Please identify and provide the rationale for any other studies or effects that should be considered.
 - a. Developmental toxicity: Archibong et al. (2002) [decreased fetal survival]

COMMENT:

Archibong et al. (2002) exposed pregnant F344 rats to BaP ($25, 75$ and $100 \mu\text{g}/\text{m}^3$) on carbon black particles via nose-only exposure from gestation day 11 through 20. Fifty-five percent of the particles were reported to be less than 2.5 microns in size. Each group contained 10 pregnant dams. Archibong et al. (2002) found that there was no effect of BaP inhalation in female F344 rats on implantation sites or crown-rump length of pups. There was a statistically significantly reduced number of pups per litter and survival

(litter %) and at the 25 $\mu\text{g}/\text{m}^3$ dose level and a reduction of pup weight (g/litter) at the 75 $\mu\text{g}/\text{m}^3$ level. Benchmark dose modeling was attempted, but it was confirmed that the data as presented (means +/- SEM) were not amenable to benchmark dose modeling. The following table summarizes the results of Archibong et al. (2002).

Summary of Results from Archibong et al. (2002)

Endpoint	NOAEL ($\mu\text{g}/\text{m}^3$)	LOAEL ($\mu\text{g}/\text{m}^3$)
Implantation Sites	100	None
Pups per litter	0	25
Survival (litter %)	0	25
Pup weight (g/litter)	25	75
Crown-rump length (mm/litter)	100	None
Plasma progesterone, day 15	75	None
Plasma progesterone, day 17	25	75
Plasma estradiol-17 beta, day 15	75	None
Plasma estradiol-17 beta, day 17	25	75
Plasma prolactin, day 15	75	None
Plasma prolactin, day 17	25	75

However, the results of an almost identical study performed more recently in the same laboratory (Wu et al., 2003) showed decreased pup survival after BaP inhalation exposure on gestation days 11-20, but in this study, the decreased survival was not seen until higher doses. The NOAEL for decreased pup survival was 25 $\mu\text{g}/\text{m}^3$ and the LOAEL was 75 $\mu\text{g}/\text{m}^3$. The Wu et al. (2003) results from the same laboratory with the same aerosol generator and nose-only inhalation chambers in the same strain of rats undermines USEPA (2013)'s selection of 25 $\mu\text{g}/\text{m}^3$ as a LOAEL for decreased pup survival. The following table summarizes the results of Wu et al. (2003).

Summary of Results from Wu et al. (2003)

Endpoint	NOAEL ($\mu\text{g}/\text{m}^3$)	LOAEL ($\mu\text{g}/\text{m}^3$)
Survival (Pups/litter) / (implantation sites/litter)	25	75

USEPA (2013) should use a NOAEL of $25 \mu\text{g}/\text{m}^3$ for decreased pup survival in deriving candidate RfDs. USEPA (2013) rejected this study because the data were presented graphically and not numerically, but this is not a reasonable exclusion criterion. The NOAEL is easily determined from the paper, and USEPA (2013) used graphically presented data from other papers, such as Chen et al. (2012).

- b. Reproductive toxicity: Archibong et al. (2008) [decreased testes weight and decreased sperm count and motility]

COMMENT: No comment due to time constraints. No comment does not mean tacit agreement.

7. The NOAEL/LOAEL approach was used to derive the PODs for the candidate values. Please comment on whether this approach is scientifically supported and clearly described.

COMMENT:

Benchmark dose modeling was attempted, but it was confirmed that the data as presented (means +/- SEM) were not amenable to benchmark dose modeling. USEPA (2013) has focused on pup survival as the critical endpoint and defined $25 \mu\text{g}/\text{m}^3$ as the LOAEL from this study. Later, a standard Uncertainty Factor for LOAEL to NOAEL of 10 is applied to derive the candidate RfD based on pup survival. Thus, the inferred NOAEL for pup survival is $2.5 \mu\text{g}/\text{m}^3$. There are many other endpoints in the study from which to derive a NOAEL or LOAEL for dose-response assessment. The commenters recommend that USEPA derive candidate RfDs from several endpoints and then average the results to define the RfD so as to take full advantage of the weight of evidence from this study upon which USEPA (2013) relies. The recommended use of the Archibong et al. (2002) and Wu et al. (2003) studies is shown in the following tables.

Point of Departure Doses from Archibong et al. (2002) for Reference Dose Derivation

Endpoint	POD adjusted ($\mu\text{g}/\text{m}^3$)	POD Type	POD _{HEC} ($\mu\text{g}/\text{m}^3$)
Pups per litter	4.2	LOAEL	4.6
Survival (litter %)	4.2	LOAEL	4.6
Pup weight (g/litter)	4.2	NOAEL	4.6
Plasma progesterone, estradiol-17 beta &	4.2	NOAEL	4.6

prolactin (day 17)			
--------------------	--	--	--

Point of Departure Doses from Wu et al. (2003) for Reference Dose Derivation

Endpoint	POD adjusted ($\mu\text{g}/\text{m}^3$)	POD Type	POD _{HEC} ($\mu\text{g}/\text{m}^3$)
Survival (Pups/litter) / (implantation sites/litter)	4.2	NOAEL	4.6

8. Please comment on the rationale for the selection of the uncertainty factors (UFs) applied to the PODs for the derivation of the candidate values. Are the UFs appropriate, based on the recommendations described in Section 4.4.5 of *A Review of the Reference Dose and Reference Concentration Processes* (USEPA, 2002), and clearly described? If changes to the selected UFs are proposed, please identify and provide scientific support for the proposed changes.

COMMENT:

The commenters agree with most of the Uncertainty Factors but disagree that all candidate RfCs require a "Database Uncertainty Factor" of 10. BaP has one of the most complete sets of toxicological data of any chemical ever studied by toxicologists. There is no need to account for "database uncertainty," and this factor of 10 should be removed from all RfC derivations.

9. From the candidate values, an organ/system-specific reference value was selected for the developmental hazard. A reference value for the reproductive hazard was not selected due to significant uncertainty in deriving the reproductive candidate values. USEPA concluded that the developmental value best represented the hazard considering the effect on fetal survival is the most sensitive noncancer developmental effect observed following inhalation exposure to benzo[a]pyrene. Please comment on whether the selection of the organ/system-specific reference value is scientifically supported, appropriate for development of a chronic RfC, and is clearly described. Please identify and provide the rationale for any other values that should be considered.

COMMENT:

The commenters disagree that the candidate RfD based on pup survival in Archibong et al. (2002) is the appropriate dose-response value from that study and also disagree that the companion study of Wu et al. (2003) from the same laboratory should be ignored. The Wu et al. (2003) study supports a NOAEL of 25 $\mu\text{g}/\text{m}^3$ for the same endpoint versus 25 $\mu\text{g}/\text{m}^3$ as a LOAEL as determined by USEPA (2013). The

commenters recommend that USEPA derive candidate RfDs from several endpoints from Archibong et al. (2002) and Wu et al. (2003) and then average the results to define the RfD so as to take full advantage of the weight of evidence from these two studies in the same laboratory.

10. The proposed overall RfC was based on decreased fetal survival observed by Archibong et al., (2002). Please comment on whether the selection of this RfC is scientifically supported and clearly described. Please identify and provide the rationale for any other values that should be considered.

COMMENT:

As discussed above, the selection of one candidate RfC based on one endpoint from the Archibong et al. (2002) study does not represent the scientific weight of evidence from the Archibong et al. (2002) study or the companion study, Wu et al. (2003) in which the LOAEL for pup survival from the first study was not a LOAEL. It was a NOAEL instead. The following table lists the recommended candidate RfDs and the average RfD that should be proposed.

Candidate Reference Concentrations for Benzo(a)pyrene

Endpoint/Reference	POD _{HED} (µg/m ³)	Type	UF _A	UF _H	UF _L	UF _S	UF _D	Composite UF	RfD (mg/m ³)*
Developmental									
Decreased pups per litter Archibong et al. (2002)	4.6	LOAEL	3	10	10	1	1	300	1.5x10 ⁻⁵
Decreased fetal survival Archibong et al. (2002)	4.6	LOAEL	3	10	10	1	1	300	1.5x10 ⁻⁵
Decreased pup weight Archibong et al. (2002)	4.6	NOAEL	3	10	1	1	1	30	1.5x10 ⁻⁴
Decreased plasma hormone levels Archibong et al. (2002)	4.6	NOAEL	3	10	1	1	1	30	1.5x10 ⁻⁴
Decreased fetal survival	4.6	NOAEL	3	10	1	1	1	30	1.5x10 ⁻⁴

Wu et al. (2003)									
Average Value									1x10 ⁻⁴

*RfD was converted to mg/m³

In conclusion, the commenters recommend that USEPA (2013) not focus on one adverse effect in the Archibong et al. (2002) study, but instead consider the scientific weight of evidence from that study and the companion study, Wu et al. (2003) to derive an average RfC of 1X10⁻⁴ mg/m³.

Cancer Risk Estimates

Oral Slope Factor (OSF)

Carcinogenicity studies examining oral exposure to benzo[a]pyrene were evaluated and the most relevant, informative studies and endpoints were selected for dose-response analysis, where data were amenable, for consideration in deriving an OSF.

11. The Kroese et al. (2001) and Beland and Culp (1998) studies were selected as the best available studies for dose-response analysis. The incidence data for forestomach and oral cavity, liver, jejunum/duodenum, kidney, and skin tumors in male and female rats reported by Kroese et al. (2001) and forestomach, esophagus, tongue, larynx tumors (alimentary tract) in female mice reported by Beland and Culp (1998) were selected for dose-response analysis. Please comment on whether the evaluation, selection, and relevance of studies and endpoints for dose-response analysis is scientifically supported and clearly described. Please identify and provide the rationale for any other studies or endpoints that should be considered.

COMMENT:

The commenters agree that the studies of Kroese et al. (2001) and Beland and Culp (1998) are the best available studies for the assessment of the carcinogenic potency of BaP in rodents. The commenters disagree, however, that forestomach tumors are relevant to the assessment of human health because humans do not have forestomachs. When this criticism is made, USEPA typically responds that its cancer assessment guidelines do not require tumor site concordance. USEPA also comments that esophageal tissue is similar in nature to rodent forestomach tissue. However, in this particular case, esophageal tumor results were observed and dose-response modeling can be performed on esophageal tumors rather than relying on a surrogate tissue. Given that USEPA normally uses rodent forestomach tumor incidence as a

surrogate for esophageal tumor risk in humans, the commenters recommend that this approximation step be omitted. Instead, USEPA (2013) should directly model the esophageal tumor incidence in rodents and use those results to make estimates of human risk in esophageal tissues.

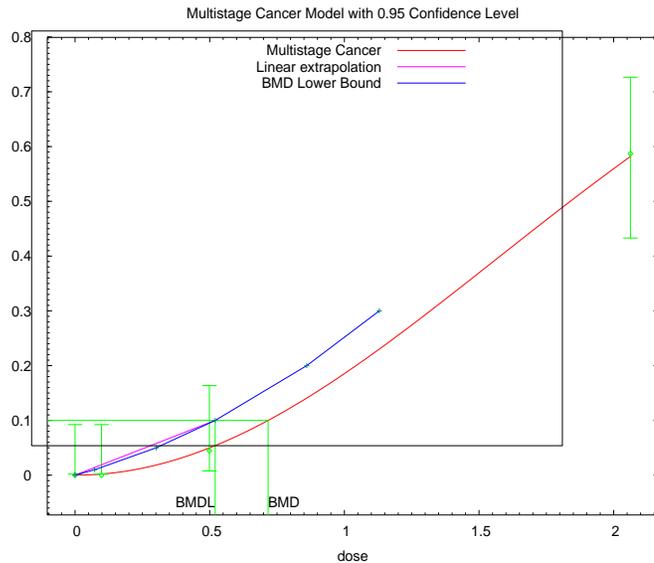
Forestomach tumor incidence data overestimates human risk. The forestomach is an organ that holds ingested food before entry into the stomach. When BaP-laced rodent food is fed to rodents, the BaP has a longer contact time with the forestomach membranes that it does with the membranes of the stomach or intestines. Humans do not have a forestomach or other organ that holds food prior to entry to the stomach, so BaP in food ingested by humans travels through the esophagus quickly before entering the stomach. The contact time of BaP in food with esophageal tissues is fast in both rodents and humans, so tumor incidence data from actual rodent esophageal tissue is a much more logical and appropriate data set for estimating human risk.

In conclusion, the commenters recommend that USEPA (2013) base the oral slope factor on esophageal tumor data, not data on the rodent forestomach.

12. BMD modeling was conducted using the incidence of the individual tumor types reported in Kroese et al. (2001) and Beland and Culp (1998) in conjunction with dosimetric adjustments for calculating the human equivalent doses to estimate the PODs. The candidate OSFs were calculated by linear extrapolation from the PODs (i.e., the lower 95% confidence limit on the dose associated with 10% extra risk). Please comment on whether this approach is scientifically supported and clearly described.

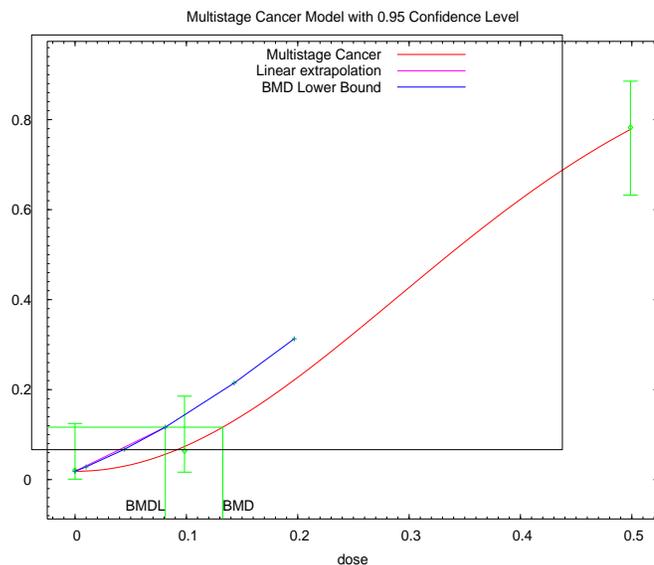
COMMENT:

USEPA (2013) should fully implement the 2005 risk assessment guidelines and not default to a linear low dose extrapolation for every OSF calculation. Benchmark dose modeling of the Beland and Culp (1998) data using the Multistage Cancer model gives plots that show evidence of a threshold for carcinogenic risk. In the case of esophageal cancer, which is the carcinogenic endpoint recommended in these comments, the lowest non-zero dose yields a 0% tumor incidence showing unequivocal evidence of a threshold.



17:17 04/07 2010

The same is true for cancer of the tongue and larynx. When forestomach tumor model plots are observed, there is also clear evidence of a threshold, and linear extrapolation overestimates the risk at low doses.



08:58 04/09 2010

13. The OSF associated with alimentary tract tumors in female mice as reported by Beland and Culp (1998) was selected as the recommended slope factor for assessing human cancer risk following oral exposure to benzo[a]pyrene. Please comment on whether this selection is scientifically supported and clearly described. Please identify and provide the rationale for any other studies or endpoints that should be selected to serve as the basis for the OSF.

COMMENT:

The commenters have performed benchmark dose modeling on the data of Beland and Culp (1998) and found a BMDL₁₀ of 0.08 mg/kg-day for forestomach tumors which is similar to the USEPA (2013) BMDL₁₀ of 0.07 mg/kg-day for total alimentary tract tumors. Both result in an estimated oral slope factor of 1 (mg/kg-day)⁻¹. However, when the esophageal tumor data (papillomas and carcinomas) was modeled, the BMDL₁₀ was 0.5 mg/kg-day and the oral slope factor was 0.2 (mg/kg-day)⁻¹. With tongue tumor data, the BMDL₁₀ was 0.6 mg/kg-day and the oral slope factor was 0.2 (mg/kg-day)⁻¹. The oral slope factor should be based on esophageal tumors with a value of 0.2 (mg/kg-day)⁻¹.

Inhalation Unit Risk (IUR)

The benzo[a]pyrene inhalation database for carcinogenicity consists of a lifetime inhalation bioassay and several intratracheal instillation studies. The instillation studies were not considered for dose-response analysis because use of this exposure method alters the deposition, clearance, and retention of substances, and therefore, is less relevant and informative for the quantitative estimation of inhalation cancer risk compared with inhalation bioassays.

14. The Thyssen et al. (1981) study was selected as the best available study for dose-response analysis as it represents the only lifetime inhalation cancer bioassay available for describing exposure-response relationships for cancer from inhaled benzo[a]pyrene. The incidence data for tumors of the upper respiratory and digestive tracts (pharynx, larynx, trachea, esophagus, nasal cavity, and forestomach) reported by Thyssen et al. (1981) were selected for dose-response analysis. Please comment on whether the evaluation, selection, and relevance of studies and endpoints for dose-response analysis is scientifically supported and clearly described. Please identify and provide the rationale for any other studies or endpoints that should be considered.

COMMENT:

The commenters agree that Thyssen et al. (1981) is an inhalation study that used inhalation of BaP aerosols. However, USEPA (2013) fails to report the study of Pauluhn et al. (1985) in the same laboratory with the same Syrian golden hamsters and the same aerosol methods. This study must be considered in the derivation of the IUR. Several other studies are available in which intratracheal or intrabronchial instillation were used as modes of administration, but such studies are clearly inappropriate for dose-response assessment. However, the Thyssen et al. (1981) study is also inappropriate for dose-response modeling and should be abandoned. The animals were dosed with BaP adsorbed onto sodium chloride aerosol. The method of creating the aerosol is complex. BaP was vaporized in a boiler and passed to a heater. The salt aerosol was created by spraying a sodium chloride solution into a boiler and a heater and then mixed with the BaP vapors. Then, the mixture was heated and then cooled where the BaP condensed onto the sodium chloride particles. The BaP in the air that the animals inhaled was determined by fluorometry.

The concentration of salt aerosol in the air is not reported. Control animals were exposed to a sodium chloride aerosol of 240 µg/m³, but the paper does not report why this value was chosen for the control

animals. Clearly the dosed animals had much higher levels of total particles in the air they breathed than the control animals, because the dosed animals received 2,200 to 46,500 $\mu\text{g}/\text{m}^3$ of BaP plus an unknown amount of salt versus the controls, which received 240 $\mu\text{g}/\text{m}^3$ of sodium chloride. The appropriate control group would have been exposed to at least 9,500 $\mu\text{g}/\text{m}^3$ of sodium chloride given that the low BaP dose group did not exhibit any respiratory tumors.

Particle overload is a problem with studies that use such high concentrations as 47 mg/m^3 BaP. For reference, the typical levels of BaP in indoor and outdoor air are measured in fractions of nanograms per cubic meter, not milligrams per cubic meter. Northcross et al. (2012) measured the BaP level in a car in which three cigarettes were smoked over 3 hours without ventilation, and the measurement was 8 ng/m^3 . Typical levels in smoker's houses are less than 1 ng/m^3 . So the experiment performed by Thyssen et al. (1981) used exposure levels that were over a *million times* higher than the levels that would reasonably be expected for humans even in environments where BaP levels are higher than usual.

The levels at which particle overload have been documented in the literature are less than 47 mg/m^3 . As reported by Oberdorster (1995), NTP (1993) concluded that particles "...the maximum ability of the respiratory tract to clear particles was exceeded at the 6 and 18 mg/m^3 exposure levels." Oberdorster (1995) recommended that the TLV for dust (nuisance dust) be lowered from 5 mg/m^3 to 1 mg/m^3 to prevent particle overload. Levy (1995) reports that several models indicate that particle concentrations of 1-2 mg/m^3 should be defined as the occupational exposure level for "particles not otherwise classified," and that the ACGIH had recently proposed a TLV of 3 mg/m^3 to protect against lung overload of particles. Clearly, the two highest dose levels used in the Thyssen et al. (1981) study were so high as to have caused particle overload, greatly overestimating the actual risk that would be relevant to human exposures at concentrations thousands or millions of times lower in concentration.

Doses that cause particle overload clearly exceed the Maximally Tolerated Dose (MTD), and USEPA guidance for carcinogenic risk assessment prohibit USEPA from using data exceeding the MTD for risk quantitative assessment purposes. USEPA (2005) states:

"Animals studies are conducted at high doses in order to provide statistical power, the highest dose being one that is minimally toxic (maximum tolerated dose or MTD). Consequently, the question often arises of whether a carcinogenic effect at the highest dose may be a consequence of cell killing with compensatory cell replication or of general physiological disruption rather than inherent carcinogenicity of the tested agent.... If adequate data demonstrate that the effects are solely the result of excessive toxicity rather than carcinogenicity of the tested agent per se, then the effects may be regarded as not appropriate to include in assessment of the potential for human carcinogenicity of the agent."

"In the case of inhalation studies with respirable particles, evidence of impairment of normal clearance of particles from the lung should be considered along with other signs of toxicity to the respiratory airways to determine whether the high exposure concentration has been appropriately selected (USEPA, 2001a)."

There is also considerable confusion about the dosage, the number of animals and the number of tumors in the Thyssen et al. (1981) study. The following table shows that there is uncertainty about the most basic

aspect of any toxicology study, which is the number of animals exposed to the test agent. This is due to the manner in which the experiment was performed. According to Thyssen et al. (1981), each group started with 24 hamsters each, but if an animal died within the first 12 months, it was replaced. It would appear from the data presented in the paper that more animals died within the first 12 months in the two groups that did not develop tumors (control and lowest BaP dose group). A study in which control animals had high mortality is not suitable for dose-response assessment.

Uncertainty about Number of Animals in Thyssen et al. (1981)

Test Group	Thyssen et al. (1981)	USEPA (1990)	USEPA (2013) Page D-54	USEPA (2013) Table D-13	USEPA (2013) Table D-14	USEPA (2013) Table E-17	USEPA (2013) BMDM
Control	27*	22	20-30	27*	27	27	23
Dose 1	27*	24	20-30	27*	27	27	24
Dose 2	26*	24	20-30	26	26	26	26
Dose 3	25*	23	20-30	25	34	33	23

*Effective number of animals; “effective” not defined

USEPA (2013) creates even more confusion and uncertainty by relying on an earlier USEPA (1990) analysis of this study and then disagreeing with their own study and with the published paper about the number of animals and disagreeing with themselves about the number of animals in five places within the document under review. Knowing the number of animals in each group is not an insignificant piece of information for dose-response modeling. For instance, the consensus seems to be that there were 27 animals in the control and first dose groups. However, USEPA’s benchmark dose modeling assumed 23 and 24 for some reason. Assuming that are less animals than there really were in the two groups that did not get tumors, arbitrarily give less weight to zero tumor incidence. This biases the slope and the benchmark to higher potency.

There is also a total lack of clarity about how many tumors were observed and what the tumor incidences were in each group of hamsters. As noted below, there is as much as a 20-30% discrepancy between the number of tumor bearing animals in the various groups. Deviating by 20-30% in tumor incidence renders any quantitative dose-response modeling *unquantitative* and plainly *wrong*.

Uncertainty about Number of Tumor-Bearing Animals in Thyssen et al. (1981)

Dose Group	USEPA (1990)	USEPA (2013) Table D-13	USEPA (2013) Table D-14	USEPA (2013) Table E-17
Dose 2 Larynx	11	8	11	11
Dose 2 Pharynx	9	6	9	9
Dose 3 Larynx	12	13	12	13
Dose 3 Pharynx	18	14	18	18

The dose given to the hamsters is also highly uncertain. According to USEPA (1990): “Data exist for exposure [sic] measured in the actual exposure chambers over the three years that the entire experiment was conducted. Variability of the measurements over time from the corresponding nominal value was apparent. Also, the duration and frequency of exposure varied among time segments of the experimental period, and animals were exposed only in segments of the entire experimental period.” It is highly unusual to expose animals for varying periods of time over the course of an experiment.

Variability in the lifetime average exposure received by each animal was high. For instance, according to USEPA (1990), the average lifetime exposure in the middle exposure chamber varied from 0.842 mg/m³ to 1.061 µg/m³. The most highly exposed hamster received a 26% higher exposure than the least exposed hamster. USEPA (2013) has acknowledged this fact: “...weekly averages of chamber concentration measurements varied two- to fivefold from the overall average for each group, which exceeds the limit for exposure variability of <20% for aerosols recommended by OECD (2009).”

USEPA (2013) cannot model a group of animals that all received different average lifetime doses, so they modeled the average of the average lifetime doses. However, most of the animals with tumors received doses higher than the average dose. Specifically, 14 of 16 (88%) tumor bearing animals in the middle dose group had a lifetime average dose of BaP higher than the average of 1.01 mg/m³ and 15 of 19 (79%) tumor bearing animals in the high dose group received a higher average lifetime dose than the average of the average lifetime doses, which was 4.29 mg/m³. Clearly, calculating a lifetime average dose for each animal considering that they were exposed for some of the time and not exposed for some of the time and then calculating a grand average for the entire dose group *overestimates* risk. The animals developing tumors did so by receiving higher actual doses than the arbitrary dose assigned to the entire group. USEPA (2013)

Table D-4 shows that the benchmark dose modeling was, indeed, performed assuming that all animals in each group received an average lifetime concentration of 0, 0.25, 1.01 or 4.29 mg/m³. This over estimation of doses leads to unrealistic estimates of risk.

It is also incorrect to assume that the only metric that matters is lifetime average dose. It is true that doses are often adjusted for lack of continuous exposure such as applying a 5/7 factor to doses in experiments in which animals are doses 5 days a week. This is a minor adjustment for an experiment in which the animals received almost continuous exposure to the test agent. In this case, "animals were only exposed in segments of the entire exposure period" (USEPA, 1990). Furthermore: "Each animal's lifetime average exposure value was obtained by considering the interval that the animal was on test, shown for a few animals in Figures 1-3. It was assumed that the exposure was zero for all time periods an animal was not in the exposure chamber" (USEPA, 1990).

NRC (2011) is asking USEPA to be more transparent about how it derives dose-response factors in the IRIS program, but the derivation of the Unit Risk for BaP using the Thyssen et al. (1981) data does not meet their recommendations. The commenters recommend that USEPA (2013) needs to abandon the effort given the poor quality of this study.

In addition, USEPA (2013) fails to report the study of Pauluhn et al. (1985) in the same laboratory with the same Syrian golden hamsters. In this two-year nose-only inhalation study of BaP (2 and 10 mg/m³) on sodium chloride coated particles, the results were reported as "...few neoplastic alternations were found." Pauluhn et al. (1985) is a one-page report of a significant study that appears to contradict the results of Thyssen et al. (1981). USEPA (1990) took the trouble to request all of the detailed data from the 1981 hamster study, but not the contradictory data from the 1985 study. Before releasing a Unit Risk for BaP using the ill-defined data from Thyssen et al. (1981), the data of Pauluhn et al. (1985) should be thoroughly evaluated.

15. BMD modeling was conducted using the overall incidence of the tumors of the upper respiratory and digestive tracts reported by Thyssen et al. (1981) to estimate the PODs. Dosimetric adjustments for calculating the human equivalent concentrations were not conducted due to the lack of data to inform a basis for extrapolation to humans. It was assumed that equal risk for all species would be associated with equal concentrations in air; thus, the continuous time-weighted group average concentrations in male hamsters were used for the dose-response analysis under the assumption that these are representative across species. The candidate IURs were calculated by linear extrapolation from the PODs (i.e., the lower 95% confidence limit on the concentration associated with 10% extra risk of tumors of the upper respiratory and digestive tracts). Please comment on whether this approach is scientifically supported and clearly described.

COMMENT:

USEPA (2013) provides a lengthy discussion of the many uncertainties associated with its proposed IUR for BaP. These include:

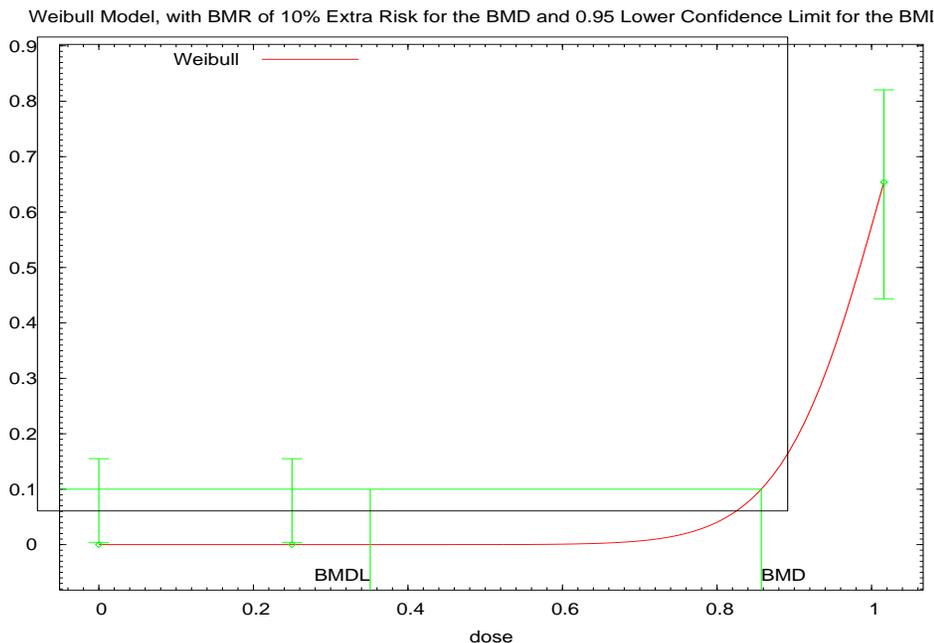
- Uncertainty and variability in the dose given to each animal
- Uncertainty in particle size distributions
- Inability to use its dosimetric model for extrapolating to hamsters to humans
- Use of a hygroscopic carrier (sodium chloride)

The inability to use an extrapolation model is because USEPA's default model is designed for insoluble and nonhygroscopic particles. Accordingly, USEPA assumed that average lifetime concentration in air yields equal risks between hamsters and humans: "This is equivalent to assuming that any metabolism of benzo[a]pyrene is directly proportional to breathing rate and that the deposition rate is equal between species."

The vast literature on BaP metabolism clearly demonstrates that metabolizing enzymes in the upper respiratory tract of rodents differs from humans. This has been studied in great detail with regard to naphthalene metabolism in rats in specific regions of the upper respiratory tract. In addition, it is well known that the upper respiratory tract morphology of rodents is very different from humans. Thus, it is not credible to assume that the risks posed by inhalation of BaP on soluble salt particles is equivalent across species. Assuming equal deposition and metabolism overestimates human risk. USEPA's detailed assessment of this critical issue is: "There are no data to support alternatives. Equal risk per $\mu\text{g}/\text{m}^3$ is assumed."

USEPA (2013) has defaulted to assuming a linear extrapolation to zero but the data of Thyssen et al. (1981) clearly supports a threshold for cancer. USEPA's benchmark dose modeling yields a benchmark concentration level (BMCL₁₀) of $0.198 \text{ mg}/\text{m}^3$ from an experiment in which animals exposed for a lifetime to an estimated $0.25 \text{ mg}/\text{m}^3$ did not develop any laryngeal or pharyngeal tumors at all. In fact, $0.25 \text{ mg}/\text{m}^3$ is a NOAEL for upper respiratory tract tumors in hamsters.

All of the plots from the Benchmark Dose Modeling Software for models giving fits using USEPA's criteria clearly show the non-linear nature of the dose-response curve. These plots are shown in Appendix A, but the plot for the Weibull model is shown here as an example.



Thus, it is not appropriate to use linear extrapolation for the data from Thyssen et al. (1981), but if it were done using the $BMCL_{10}$ from USEPA's benchmark dose modeling, the IUR should be based on the $BMCL_{10}$ from the model runs where tumors were considered the cause of death.

Furthermore, the commenters were unable to re-create the USEPA (2013) reported $BMCL_{10}$ values. When the data were modeled using USEPA's Benchmark Dose Modeling Software, no model fits were found. When the Software was run with the highest dose removed, there were nine good model fits. The average $BMCL_{10}$ was 0.35 mg/m^3 .

16. The IUR associated with tumors of the upper respiratory and digestive tracts in male hamsters (in which the tumors were considered incidental to the death of an animal) as reported by Thyssen et al. (1981) was selected as the recommended unit risk for assessing human cancer risk following inhalation exposure to benzo[a]pyrene. Please comment on whether this selection is scientifically supported and clearly described. Please identify and provide the rationale for any other studies or endpoints that should be selected to serve as the basis for the IUR.

COMMENT:

USEPA (2013) states that it has no knowledge one way or another whether deaths in the hamsters were related to tumors or not. Thyssen et al. (1981) clearly report that survival was greatly reduced from 96%, 95% and 96% in the first three groups to 60% in the high dose group. Given the high mortality rates reported, it is more reasonable to assume that the tumors contributed to the cause of death of this large number of animals. As noted above, it is not appropriate to use linear extrapolation for the data from Thyssen et al. (1981), but if it were done using the $BMCL_{10}$ from USEPA's benchmark dose modeling, the

IUR should be based on the BMCL₁₀ from the model runs where tumors were considered the cause of death. The BMCL₁₀ would be 0.461 mg/m³ and the IUR would be 2 x 10⁻⁴ (µg/m³)⁻¹.

Dermal Slope Factor (DSF)

Carcinogenicity studies examining dermal exposure to benzo[a]pyrene were evaluated and the most relevant, informative studies and endpoints were selected for dose-response analysis, where data were amenable, for consideration in deriving a DSF.

17. The Roe et al. (1970), Sivak et al. (1997), and Poel (1959) studies were selected as the best available studies for dose-response analysis. Several other studies provided supportive information but were considered less informative due to incomplete exposure duration information or greater uncertainty associated with extrapolating to lower doses. These studies were included in the dose-response analysis to help characterize similarities among the studies on a quantitative basis. The incidence data for skin tumors in male and female mice were selected for dose-response analysis. Please comment on whether the evaluation, selection, and relevance of studies and endpoints for dose-response analysis is scientifically supported and clearly described. Please identify and provide the rationale for any other studies and endpoints that should be considered.

COMMENT:

A dermal slope factor (DSF) should not be derived at all for the many reasons discussed above. In this comment, the details of the DSF are discussed. The above listed studies are not the best available studies. Several studies were omitted without cause, and several studies were *considered less informative* because of erroneous characterizations of the studies.

Dermal Dosimetry Not Amenable to Dose-Response Assessment

Dermally administered BaP dose is not cleared quickly as are oral doses. Instead, the BaP builds up, so the whole concept of “daily dose” is erroneous. As each dosing goes by, the amount of BaP in the skin is building up as a “depot” or “reservoir” dose. The average daily dose is irrelevant and inappropriate for dermal dose-response assessment.

For instance, Fasano and McDougal (2008) measured the dermal absorption rate of 34 chemicals and counted the amount of chemical in skin after the end of the experiment as “absorbed” chemical because of the known depot effect.

Kao et al. (1988) studied BaP penetration in mouse skin in seven different mouse strains in an *in vitro* system. After 16 to 18 hours, almost all of the dose of applied BaP was recovered. Most of the dose was present in the skin digest rather than in the receptor fluid showing that it was present within the matrix of the skin, forming a “depot” dose.

Moody et al. (2007) performed studies in human skin with BaP in an *in vitro* system. He found that after 24 hours, the BaP in acetone, 56% of the applied dose, was not recovered by soap and water wash and counted 56% of the dose as “absorbed.” However, 80% of that dose was contained in the skin “depot” and had not yet penetrated the skin. In a 42-hour study, 50% was not recovered by a soap and water wash and was counted as “absorbed.” As above, 78% of that dose was contained in the skin “depot.”

Researchers performing dermal penetration studies routinely score the amount of chemical that cannot be removed from the skin by a soap and water wash as “dedicated” dose that is considered “absorbed.” This is a conservative stance that assumes that even if the chemical has not penetrated the skin by the time the dermal experiment was completed, if it had entered the skin deep enough that it could not be easily removed, then it should be counted as “absorbed.” The implications for dose-response assessment, however, are that repeated doses will increase the depot dose. Thus, the true daily skin dose is much higher than the daily *administered* dose.

Specifically, OECD (2004) is its *OECD Guideline for the Testing of Chemicals: Skin Absorption: in vitro Method* states the following: “The absorbable dose (in vitro) represents that present on or in the skin following washing.”

Further, OECD (2010) in its *OECD Guidance Notes on Dermal Absorption*, stated:

“A study duration of more than 24 hours is not recommended because skin tissue can be expected to deteriorate. Of course, for some substances, in particular those that are lipophilic, it may take longer for a chemical to migrate from a skin depot to the receptor fluid. From a regulatory point of view, however, the resulting uncertainty can be readily overcome by including the amount found in the skin as potentially absorbed...”

“Dermal absorption is primarily a diffusion-driven process, and therefore test substance in the lower layers of the *stratum corneum* should be assumed to form a reservoir that may become systemically available, unless it can be demonstrated *in vivo* that absorption is complete and this test substance will remain in the *stratum corneum* until exfoliated (see 7.1.3).”

“The question of whether to include skin-bound residues is addressed in Section 7.1. For *in vitro* studies, the OECD guideline (OECD 428) defines the ‘absorbable dose’ as ‘that present on or in the skin following washing’. A similar approach is recommended for the *in vivo* studies.”

In addition to the general information from the dermal penetration literature on PAHs, a Key Study from USEPA (2013) directly addresses this issue. Poel (1959) measured BaP fluorescence on the mouse skin and found that the BaP persists from one dosing event to another.

“In our own experiments, fluorescence of the exposed skin has been observed to persist for more than a week after a single application of benzopyrene, and comparable fluorescent periods have been observed with the more potent agents, 9,10-dimethyl-1,2-benzanthracene (DMBA) and methyl cholanthrene (MCA),

on mouse skin. Apparently, the ‘single exposures’ of past experiments were in effect exposures of extended duration.”

Thus, the actual doses the skin received in this and all of the studies cannot be converted to an *average daily dose* by simply taking the doses given twice or three times a week and averaging them over seven days as is typical for oral dosing experiments. The doses, instead, persisted and concentrated in the skin as each new dose was given. So the actual doses to the skin were much, much greater than those modeled by USEPA, rendering the entire concept of dose-response modeling from mouse skin studies impossible and USEPA’s proposed DSF incorrect.

Lastly, USEPA’s *Benchmark Dose Technical Guidance Document* (USEPA, 2000a) states: “The data set should contain information relevant to dose-response for modeling.” None of the Key Studies provide *relevant* data for dose-response modeling because they ignore the true skin dose at any given point in time.

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)

The exclusion criterion is discussed below. More importantly, none of the Key Studies is suitable for dose-response assessment. Reasons for their unsuitability for dose-response assessment include:

- Inadequate and Poorly Defined Dosimetry
- Exceedance of Maximum Tolerated Dose

Study Duration

USEPA (2013) chose to perform dose-response modeling only on mouse skin studies that were greater in duration than 52 weeks. The unstated rationale for this is most likely that cancer feeding studies are usually two years in duration. USEPA (2013) incorrectly assumed this to be the case for mouse skin studies. In addition, certain mouse skin studies were designed to be executed for a full 104 weeks *or until the animals died*. However, the high doses used in many studies resulted in early mortality leading to shorter study durations.

For instance, Levin et al. (1977) was dismissed because the study duration was presumably less than 52 weeks. However, this study was actually 60 weeks in duration. Nesnow et al. (1983), a study performed by

USEPA scientists, was also dismissed because it was stated to have been less than a year. In fact, Nesnow et al. (1983) dosed their mice for 50-52 weeks. These two studies and other studies that were excluded from dose-response modeling should, instead, be included because 104-weeks are not required to observe skin tumors in mice. As noted in the table below, tumors were seen by 21 to 53 weeks in the studies cited and used by USEPA (2013).

Summary of Timing of First Tumor Appearance

Study Cited in USEPA (2013)	Average Daily Dose (µg/day)	First Appearance of Tumors (weeks)
Roe et al. (1970)	0.04	29
Poel (1959)	0.06	42
Poel (1959)	0.16	24
Poel (1959)	0.32	36
Poel (1959)	1.63	21
Sivak et al. (1997)	1.4	43
Grimmer et al. (1983)	1.1	41 (mean – 2 SD)
Grimmer et al. (1984)	0.97	53 (lower 95 th CI)

In addition, many animals died in many of the USEPA-cited studies far earlier than one year, as noted in the table below. In the Poel (1959) study, the early mortality is significant. The range of survival in the control animals was 29 to 92 weeks with a median survival of 60 weeks. In the highest dose group modeled by USEPA (2013), the lifespan ranged from 25 to 82 weeks, with a median value of 56 weeks. So, about half of the animals died before one year, and many died by six months.

Summary of Survival Data from Key Studies

Study Cited in USEPA (2013)	Survival Time (weeks)	Mortality Rate (%)
Sivak et al. (1997)	90 in low dose (mean) 95 in medium dose (mean) 64 in high dose (mean)	

Poel (1959)	60 in controls (median) 61 in 0.15 µg group (median) 65 in 0.38 µg group (median) 56 in 0.75 µg group (median) 56 in 3.8 µg group (median) 39 in 19 µg group (median) 28 in 94 µg group (median) 21 in 188 µg group (median) 23 in 376 µg group (median) 23 in 752 µg group (median) 29 in controls (minimum) 16 in 0.15 µg group (minimum) 19 in 0.38 µg group (minimum) 15 in 0.75 µg group (minimum) 25 in 3.8 µg group (minimum) 21 in 19 µg group (minimum) 19 in 94 µg group (minimum) 6 in 188 µg group (minimum) 15 in 376 µg group (minimum) 13 in 752 µg group (minimum)	
Habs (1984)	99 in control animals 93 for dose level 1 57 for dose level 2	
Roe, et al. (1970)		Dose 1 16% at 43 weeks Dose 2 16% at 43 weeks Dose 3 14% at 43 weeks Dose 4 18% at 43 weeks Dose 5 20% at 43 weeks Dose 1 30% at 57 weeks Dose 2 26% at 57 weeks Dose 3 26% at 57 weeks Dose 4 26% at 57 weeks Dose 5 36 % at 57 weeks
Schmahl et al. (1977)		Control 18% (unspecified time) Dose 1 23% (unspecified time) Dose 2 12% (unspecified time) Dose 3 19% (unspecified time)

Inadequate and Poorly Defined Dosimetry

Almost all of the studies cited by USEPA (2013) fail to meet the minimal standards for selection of a study for use in dose-response modeling, because one cannot determine the actual dose or the effective dose to the rodent skin. In addition, most of the studies are old and methods to determine and transfer dose were crude.

Sivak et al. (1997): The source and purity of BaP was not discussed. Each application consisted of BaP dissolved in 50 microliters (µL) in solvent, but there is no discussion of how the 50 µL was delivered and how precisely the delivery device delivered 50 µL.

The study does not mention the surface area of the dosed area or if the mouse skin was occluded or not. The lack of information on the surface area is troubling, for example, because 1 µg/BaP onto 1 cm² is a very different effective dose than is 1 µg/10 cm².

Given that this study was performed in 1997 and did not use “one drop” as their unit of dosing, one can surmise that the transfer of 50 µL was performed with a micropipette with reasonable precision. However, the surface area is totally unknown and no details are provided on the verification of the concentrations of standard BaP solutions over the study period as is required in NTP cancer bioassays.

Poel (1959): The source and purity of BaP was not discussed. Each application consisted of a drop of one of the prepared solutions applied to the shaved interscapular skin with a blunted 20-gauge needle dropper with an assumed transfer rate of 0.0075 mL per drop. For details of dosing, the reader is referred to another Poel (1959) study. In this earlier publication, the authors reported that drops were calibrated with a syringe to deliver 127 +/- 2.4 drops per mL. Thus, delivered doses were 0.008 mL/drop.

The study does not mention the surface area of the dosed area or if the mouse skin was occluded or not. The lack of information on the surface area is troubling, because 1 µg BaP onto 1 cm² is a very different effective dose than is 1 µg/10 cm².

This 1950-vintage procedure is extremely *unquantitative* in nature, and it is a certainty that the doses varied widely over the many applications that occurred as the weeks went by. This study was typical of those carried out in the 1950s and 1960s. Such a study had and has value for hazard identification purposes, but it has no value for quantitative dose-response assessment.

Roe et al. (1970): This study presented no information whatsoever on BaP purity, concentration verification, volume of dose, surface area of dose, or any other details that would be required to qualify a study as suitable for dose-response assessment. The study says that “calibrated pipettes” were used.

Schmidt et al. (1973): The only information on dosing is that the dose was delivered as a single drop to the skin. No information was presented on the mode of delivery, concentration verification, or surface area of dose.

Schmahl et al. (1977): This study stated that the doses of BaP were delivered in a 20 µL solution with a syringe. No information was presented on BaP purity, concentration verification, surface area of dose, or any other details that would be required to qualify a study as suitable for dose-response assessment.

Habs et al. (1980): This paper reported that the dose was delivered as a solution in acetone of 20 µL with a calibrated Hamilton syringe. No information was presented on concentration verification or surface area of dose.

Habs et al. (1984): This study reported that BaP was >96% with a source provided. The dose was delivered as a solution in acetone of 10 µL with a calibrated Hamilton syringe. No information was presented on concentration verification or surface area of dose.

Grimmer et al. (1983): Grimmer et al. (1983) reported that the BaP was a solution in 100 µL acetone: DMSO (1:3). No information was presented on the mode of delivery, concentration verification, or surface area of dose.

Grimmer et al. (1984): USEPA (2013) cites Grimmer et al. (1984) as a source of data, but the citation is incorrect. The cited paper does not concern a mouse skin painting study. USEPA's data attributed to Grimmer et al. (1984) actually comes from Grimmer et al. (1985) (*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*). This paper reported that the BaP was a solution in 100 µL acetone: DMSO (1:3). No information was presented on the mode of delivery, concentration verification, or surface area of dose.

In conclusion, the doses are highly uncertain in all of these studies with no information whatsoever that would allow someone to determine if the entire dose in µg BaP was delivered to a small, medium, or large area of skin. No information was provided in any report that stock BaP solutions were prepared with some frequency or that the concentration did not increase over time as volatile solvents evaporated as the solutions were stored. None of these studies meets the minimum standards for a study suitable for dose-response assessment.

Exceedance of Maximally Tolerated Dose

Many of the studies were carried out at doses that exceeded the maximally tolerated dose (MTD). In accordance with USEPA (2005) policy, data from experiments performed above the MTD should be excluded from quantitative dose-response assessment. Specifically, USEPA (2005) states:

*"In general, while effects seen at the highest dose tested are assumed to be appropriate for assessment, it is necessary that the experimental conditions be scrutinized. Animal studies are conducted at high doses in order to provide statistical power, the highest dose being one that is minimally toxic (maximum tolerated dose or MTD). ...If adequate data demonstrate that the effects are solely the result of excessive toxicity rather than carcinogenicity of the tested agent *per se*, then the effects may be regarded as not appropriate to include in assessment of the potential for human carcinogenicity of the agent. This is a matter of expert judgment, with consideration given to all of the data available about the agent, including effects in other toxicity studies, structure-activity relationships, and effects on growth control and differentiation."*

The Office of Science and Technology Policy (1985) also discuss the MTD:

"The highest dose currently recommended is that which, when given for the duration of the chronic study, is just high enough to elicit signs of minimal toxicity without significantly altering the animal's normal life span due to effects other than carcinogenicity (19). This dose, sometimes called the maximum tolerated dose (MTD), is estimated in a subchronic study (usually of 90 days duration) primarily on the basis of adverse pathology signs, toxicity, mortality, and pathology criteria. The MTD should not produce toxicity of a severity that would interfere with the interpretation of the study. Nor should it comprise so large a fraction of the animal's diet that the nutritional composition of the diet is altered, leading to nutritional imbalance."

Further, for dermal carcinogenesis studies, USEPA (1988b) has defined the Maximum Tolerated Dose as a dose that does not cause a “marked inflammatory response or ulcerative lesion.” USEPA (1988b) specifically stated:

“It was recommended that a dose level that incites a marked inflammatory response of ulcerative lesion that is clearly related to application of the compound should not be used for an MTD.”

“Microscopic lesions of inflammation, spongiosis, degeneration, dermal edema, and possibly others, must be evaluated carefully in the selection of the MTD. If, in the opinion of the pathologist, the severity of such lesions might lead to destruction of the functional integrity of the epidermis, these lesions would indicate selecting a lower dose for the MTD.”

Clearly, mortality is not consistent with the MTD. In a few of the studies, mortality was very high. USEPA should have excluded at least some of the studies with high mortality, especially when the elevated mortality was early in the experiment. Instead of excluding datasets with high mortality, USEPA (2013) biased the tumor incidence high by decreasing the size of the animal group to include only animals that were alive at the time of the first tumor. If a small fraction of animals died for reasons unrelated to PAH administration, then this would be a reasonable statistical approach. However, when mortality is significant compared to control group mortality or is unusually high, the data should be excluded entirely from use in deriving DSFs.

In most cases, despite the fact that animals were dying, the authors of most of the USEPA-chosen studies did not make any statements about skin toxicity. The exception is Poel (1959) and Sivak et al.(1997), which are the two studies that USEPA has chosen for derivation of the DSF.

Sivak et al. (1997) reported that skin lesions were seen: “With respect to skin lesions, Group 24, with the highest dose of BaP (0.01%) applied repeatedly, had an incidence of 80% of scabs and sores.”

Poel (1959) stated that they examined animals for skin lesions, but no observations were reported. However, irritation is posited as a causal mechanism for skin tumorigenesis: “Carcinogenesis is an extreme form of reactive hyperplasia to a persistent, physiologically irreparable state of tissue damage or homeostatic imbalance, resulting ultimately in malignant overgrowth of cells related to the impaired tissue through contiguity or homeostatic mechanisms.” This statement strongly implies that chronic irritation was seen in the Poel (1959) study.

In conclusion, the doses used by Poel (1959) and Sivak (1997) (1.6 µg/day and 1.4 µg/day) were extremely high and caused significant toxicity, thus exceeding the MTD. With the exception of Schmidt et al. (1973) and Schmahl et al. (1977), all of the other studies for which USEPA performed benchmark dose modeling use similar or higher doses as seen in the table below. Thus, it is highly likely that the high dose in all studies exceeded the MTD for skin toxicity.

Summary of Highest Doses in Key Studies

Study	Highest Dose(s) (µg/day)
--------------	---------------------------------

Study	Highest Dose(s) (µg/day)
Sivak et al. (1997)	1.4
Poel (1959)	1.6
Poel (1960) (SWR)	1.6
Poel (1960) (C2HeB)	1.6
Poel (1960) (A/He)	1.6
Roe et al. (1970)	1.3, 3.9
Schmidt et al. (1973) (Swiss)	0.6
Schmidt et al. (1973) (NMRI)	0.6
Schmahl et al. (1977)	0.9
Habs et al. (1980)	1.3
Habs et al. (1984)	1.1
Grimmer et al. (1983)	2.2, 4.4
Grimmer et al. (1984)	1.9, 3.9

Few of the investigators made detailed observations about skin irritation and skin toxicity. It is likely that they were unconcerned about documenting chronic skin irritation precisely because they expected it. Hundreds of mouse skin studies with PAHs and other chemicals have been performed using the two-stage protocol for mouse skin carcinogenesis. In this model, a single dose of a test chemical is administered and then daily applications of 12-O-tetradecanolyphorbol-13-acetate (TPA) are given to *cause* chronic skin irritation. Studies in which chronic skin irritation occurs may be relevant for hazard identification, but they are not relevant for dose-response assessment. For such a purpose, studies must be performed that use doses far below the MTD and exhibit no chronic skin lesions, like the “scabs and sores” seen in USEPA’s key study (Sivak et al., 1997).

The National Toxicology Program (NTP, 1996) performed complete carcinogenesis studies of the carcinogenic PAH DMBA (2.5 µg/week) in male and female B6C3F1, Swiss (CD-1) and SENCAR mice for 52 weeks. Observation of skin lesions was a major aspect of the NTP protocol, because “chemicals with promotion potential have been reported to cause inflammation and epidermal hyperplasia.” Little irritation or ulceration was seen in the B6C3F1 mice of either sex, but >20% incidence of irritation or ulceration was seen in the other two strains at 41 to 50 weeks.

Summary: None of the Key Studies meets the minimum criteria for defining a study upon which to base dose-response modeling. USEPA (2003) in its *A Summary of General Assessment Factors for Evaluating the Quality of Scientific and Technical Information* states:

“When evaluating the quality and relevance of scientific and technical information, the considerations that the Agency typically takes into account can be characterized by five general assessment factors, including soundness, applicability and utility, clarity and completeness, uncertainty and variability, and evaluation and review. Regarding chemical toxicity testing and human health risk assessment, USEPA generally evaluates information by weighing considerations that fit within these five assessment factors.

- Soundness is defined as the extent to which the scientific and technical procedures, measures, methods, or models employed to generate the information are reasonable for, and consistent with, the intended application.

- Applicability and utility are defined as the extent to which the information is relevant for the Agency's intended use.
- Clarity and completeness are defined as the degree of clarity and completeness with which the data, assumptions, methods, quality assurance, sponsoring organizations and analyses employed to generate the information are documented.
- Uncertainty and variability are defined as the extent to which the variability and uncertainty (quantitative and qualitative) in the information or the procedures, measures, methods or models are evaluated and characterized.
- Evaluation and review are defined as the extent of independent verification, validation and peer review of the information or of the procedures, measures, methods or models."

The following table summarizes key information from the key studies that demonstrates that the key studies are all inadequately performed and reported and cannot be relied upon for dose-response assessment.

Summary of Key Studies

Study Citation	BaP source or purity defined?	BaP concentration verified?	Delivered dose quantified?	Skin surface area specified?	Exceeds MTD?	Doses averaged over dead animals?
Sivak et al. (1997)	No	No	No (no details)	No	Noted	Yes
Poel (1959)	No	No	No (one drop)	No	Noted	Yes
Poel (1960)* (SWR)	No	No	No (one drop)	No	Likely	Not known
Poel (1960)* (C2HeB)	No	No	No (one drop)	No	Likely	Not known
Poel (1960)* (A/He)	No	No	No (one drop)	No	Likely	Not known
Roe et al. (1970)	Source identified	No	Yes (calibrated pipette)	No	Likely	No
Schmidt et al. (1973) (Swiss)	No	No	No (one drop)	No	Not likely	No
Schmidt et al. (1973) (NMRI)	No	No	No (one drop)	No	Not likely	No
Schmahl et al. (1977)	No	No	Possibly (syringe)	No	Not likely	No
Habs et al. (1980)	No	No	Yes (calibrated Hamilton syringe)	No	Likely	Yes
Habs et al. (1984)	Yes (>96% purity)	No	Yes (calibrated Hamilton syringe)	No	Likely	No
Grimmer et al. (1983)	No	No	No	No	Likely	No
Grimmer et al. (1984)	No	No	No	No	Likely	No

*Actually Poel (1963)

Critique of USEPA (2013) Dismissal of Certain Studies

USEPA (2013) has dismissed certain of the studies as being “considered less informative due to incomplete exposure duration information or greater uncertainty associated with extrapolating to lower doses.” The following table listed the reasons USEPA (2013) rejects the studies for DSF derivation and presents the *actual facts*.

Summary of Reasons USEPA (2013) used to dismiss Certain Studies

Study Citation	USEPA (2013) Reasons for Dismissal	Actual Facts from the Cited Studies
Poel (1960)* (SWR mice)	No characterization of survival/exposure duration	Such information is not necessary. Average daily dose was provided. Animals were treated “until they died or a persistent skin tumor developed.” Range and median time-to-tumor was reported for each dose group.
Poel (1960)* (C2HeB mice)	No characterization of survival/exposure duration	Such information is not necessary. Average daily dose was provided. Animals were treated “until they died or a persistent skin tumor developed.” Range and median time-to-tumor was reported for each dose group.
Poel (1960)* (A/He mice)	Not listed at all	Not listed at all
Schmidt et al. (1973) (Swiss)	No characterization of exposure duration.	Such information is not necessary. Daily dose information was provided. Animals were treated until “spontaneous death of after sacrifice when neoplasms appeared.”
Schmidt et al. (1973) (NMRI)	No characterization of exposure duration.	Such information is not necessary. Daily dose information was provided. Animals were treated until “spontaneous death of after sacrifice when neoplasms appeared.”
Schmahl et al. (1977) (NMRI)	No characterization of exposure duration.	BaP was administered “until their natural death, unless they developed a carcinoma at the site of application, at which time they were killed.” Average daily dose provided.
Habs et al. (1980) (NMRI)	Higher overall exposure range; unclear overall duration of exposure	Exposure was for the animals’ lifetime for all dose groups. Survival data shown for all dose groups.

Study Citation	USEPA (2013) Reasons for Dismissal	Actual Facts from the Cited Studies
Habs et al. (1984) (NMRI)	No characterization of exposure duration for high exposure; high response a lowest exposure limits usefulness of low-dose extrapolation.	Exposure reported as “for life” and survival time given for all dose groups (648 days for low dose and 528 days for high dose); low dose gave the lowest response.
Grimmer et al. (1983) (CFLP)	No characterization of exposure duration.	Exposure duration reported as 104 weeks.
Grimmer et al. (1984) (CFLP)	No characterization of exposure duration.	Exposure duration reported as 104 weeks.

In conclusion, USEPA (2013) has focused on two studies as the “best available studies for dose-response analysis,” but these two studies are the *worst available studies for dose-response analysis* because the dosimetry was not quantitative and because Maximally Tolerated Doses were clearly exceeded. In addition, they dismiss the other available *listed* studies because they claim information is not available that clearly *is* available. Lastly, USEPA (2013) erroneously rejects for consideration two studies that are of similar or superior quality to all of the Key Studies (Levin et al. [1977] and Nesnow et al. [1983]).

18. BMD modeling was conducted using the incidence of skin tumors reported in the chronic mouse bioassays to estimate the PODs. The candidate DSFs were calculated by linear extrapolation from the PODs (i.e., the lower 95% confidence limit on the concentration associated with 10% extra risk of skin tumors). Please comment on whether this approach is scientifically supported and clearly described.

COMMENT:

USEPA (2013) has made many errors in BMD modeling. These comments have demonstrated that dose-response modeling of repeated dose mouse skin studies has no quantitative significance for human health risk assessment and that a dermal slope factor should not be derived. However, given that USEPA (2013) has chosen Key Studies and performed dose-response modeling, the specific details of USEPA’s quantitative dose-response modeling are evaluated below.

USEPA (2013) states that they used their Benchmark Dose Software (BDMS) and, in fact, that “all models available in USEPA’s Benchmark Dose Software (BDMS) were evaluated.” However, this is not an accurate statement. In the Supplemental Information document, USEPA (2013) states: “Except where other software is noted, all endpoints were modeled using the U.S. USEPA’s Benchmark Dose Software...” However, for the DSF, USEPA (2013) did not use “all models available.” Instead, they selectively used the BDMS as noted: “For each endpoint, multistage models [BMDS; (U.S. USEPA, 2012a); v 2.1] were fitted to the data using the maximum likelihood method.” So, USEPA (2013) ran *selected* models, which were only multistage models.

To verify and validate USEPA’s modeling, BDMS version 2.4 (Build 4/1/2013) was run with the data as *modified* from the original reports by USEPA (2013) and also with data as reported in the published papers.

Each key study is discussed below, and results from USEPA's BDMS (2013) are compared side-by-side with the modeling results presented in Table 2-11 and Table ES-23.

For all studies, administered doses were converted to average daily doses using the equation:

Average daily dose/day = ($\mu\text{g}/\text{application}$) \times (number of applications/week \div 7 days/week)

This is standard practice for oral dosing experiments where animals are dosed for 5 days per week and averaged by 5/7 to calculate an estimated "average daily dose." This practice is scientifically incorrect for dermal administration of PAHs which are known to be sequestered in skin and form "depot" doses. This issue has been discussed above. It invalidates the entire dose-response modeling effort performed in USEPA (2013). In fact, there is no way to simply estimate the skin dose of BaP when administered two or three times a week in an organic solvent for 52 to 104 weeks. However, in this section of comments, the details of USEPA's dose-response modeling are being addressed. .

There are two major issues that invalidate the specific details of USEPA's dose-response modeling. Both are discussed below:

- Data Adjustment
- Critical Value for Goodness of Fit

Data Adjustment

A major "data adjustment" was used in many cases. Lifetime equivalent doses were estimated for study groups that were reported to end before 104 weeks by multiplying the relevant average daily doses by $(L_e/104)^3$, where L_e is the length of exposure. Note that exposure periods <52 weeks would lead to a relatively large adjustment [i.e., $(52/104)^3 = 0.125$, or an eightfold lower *adjusted* dose than the administered dose], reflecting considerable uncertainty in lifetime equivalent dose estimates generated from relatively short studies. This adjustment was relevant for all dose groups in Poel (1959) and Roe et al. (1970), and the highest dose groups in Habs et al. (1980) and in Sivak et al. (1997).

The USEPA's *dose adjustment* is nonsensical. Take the Poel (1959) paper as an example. In this paper, the C57L mice did not live to be 104 weeks old as USEPA required. In fact, the known lifespan of the male C57L mouse is ~68 weeks (<http://www.informatics.jax.org/external/festing/mouse/docs/C57L.shtml>). In Poel's laboratory, the control mice had a median lifespan of 60 weeks. USEPA's approach to dose-response modeling was to take the average daily dose that was given to them over their lifetime and then average it over 104 weeks to determine what the dose *would* have been had these animals lived longer than they actually do. It is hard to even guess what the rationale for the *adjustment* was. The practical implications, however, are to arbitrarily reduce the actual dose to a smaller dose, that then will be modeled to a lower benchmark dose and a higher slope factor.

Summary of Erroneous Dose Adjustments in USEPA (2013)

Average Daily Dose (µg/day)	USEPA's Dose Adjustment (µg/day)
0	0
0.15	0.05
0.38	0.16
0.75	0.24
3.8	0.80

Similarly, the Roe et al. (1970) experiment was carried out for 93 weeks. The scientists stopped the experiment at 93 weeks because of high mortality in the absence of tumors. Figure 1 of Roe et al. (1970) shows an actuarial survival curve. The % that would have survived in the absence of neoplasms was 90% at week 50, 80% at week 70, 70% at week 85, and 60% at week 90. The animals did not live to 104 weeks, so there is no reasonable scientific explanation for USEPA (2013) to take the average daily dose over the lifetime of the animals and pro-rate that over 104 weeks.

The lifetime average *dose adjustment* used by USEPA is an arbitrary, nonsensical step that erroneously lowers the dose to obtain higher dermal slope factors. The dose response modeling done with *actual* average daily doses was performed for these studies by the commenters, and the results are tabularly presented below.

Critical Value for Goodness of Fit

According to USEPA's (2000a) *Benchmark Dose Technical Guidance Document*, the value to be used to determine if a model fits the empirical data well, is an alpha value of 0.1, so models with $P > 0.1$ are deemed to have adequate fits to the data. Specifically, USEPA (2000a) states:

"Since it is particularly important that the data be adequately modeled for BMD calculation, it is recommended that $\alpha = 0.1$ be used to compute the critical value for goodness of fit, instead of the more conventional values of 0.05 or 0.01."

"The guidance recommends that $\alpha = 0.1$ be used to compute the critical value for goodness of fit, instead of the more conventional values of 0.05 or 0.01, and that a graphical display of the model fit be examined as well."

USEPA (2013), on the other hand, has chosen without explanation to deviate from USEPA guidance for benchmark dose modeling and instead to use a more lenient alpha value of 0.05 as the critical value for goodness of fit. This is not appropriate and is not consistent with USEPA policy.

Interestingly, USEPA (2013) followed the *Benchmark Dose Technical Guidance Document* when modeling data for the oral RfD as noted:

"...models for the mean response were tested for adequacy of fit using a likelihood ratio test (BMDS Test 4, with χ^2 p-value < 0.10 indicating inadequate fit)." (USEPA (2013), page E-1)

So, for modeling the oral RfD, USEPA (2013) states that models with a Chi^2 p-value <0.1 are an *inadequate fit*, but for modeling the dermal slope factor, USEPA chose to ignore that guidance and, instead, conclude that models with Chi^2 p-values between 0.05 and 0.1 to be adequate fits. This is inconsistent and incorrect.

BMDS Results

Benchmark dose modeling was performed several ways to determine the implications of the above policy decisions on the BMDL_{10} determination. Each key study is discussed below.

Sivak et al. (1987): USEPA (2013) has averaged the daily dose over the potential lifetime of the animals despite the fact that the animals in the highest dose group died by 74 weeks instead of 104 weeks for the lower dose group animals. As stated above, it makes no sense scientifically to average the dose over a time period when animals are not alive and receiving chemical dosing. Doing so, in effect, artificially reduces the dose for modeling and results in a higher BMDL_{10} and DSF.

The BMDS was able to recreate USEPA's BMDL_{10} from the Multistage 2 model when the erroneous *dose adjustment* and critical value are used, but the data do not fit this model any better than they fit the other models. The average value does not differ much from the value reported in USEPA (2013). When using the proper average daily dose, the BMDL_{10} is 0.076 $\mu\text{g}/\text{day}$.

Sivak et al. (1997) Total Tumor Data	USEPA BMDL_{10} ($\mu\text{g}/\text{day}$)	Correct BMDL_{10} ($\mu\text{g}/\text{day}$)
Lifetime average dose; $P>0.05$ to define adequate fit	0.058 (Multistage 2)	0.060*
Lifetime average dose; $P>0.1$ to define adequate fit	0.058 (Multistage 2)	0.060*
Average daily dose; $P>0.1$ to define adequate fit	Not done	0.076 (LogProbit)

* Average of four BMDLs having $P>0.1$ and AICs of 49 (Multistage 2, Multistage 3, Multistage-Cancer 2, Multistage-Cancer 3)

As is typical, USEPA (2013) has merged benign and malignant tumors. However, the malignant tumor incidence was lower, and the BMDS was executed to determine the BMDL_{10} for malignant tumors. The value is 0.073 -0.082 $\mu\text{g}/\text{day}$.

Sivak et al. (1997) Malignant Tumor Data	USEPA BMDL_{10} ($\mu\text{g}/\text{day}$)	Correct BMDL_{10} ($\mu\text{g}/\text{day}$)
Lifetime average dose; $P>0.1$ to define adequate fit	Not done	0.082 (Multistage Cancer 2)
Average daily dose; $P>0.1$ to define adequate fit	Not done	0.073*

* Average of nine BMDLs having $P>0.1$ and AICs of 44 or lower (Gamma, LogLogistic, LogProbit, Multistage 2, Multistage 3, Multistage-Cancer 2, Multistage-Cancer 3, Weibull, Quantal-Linear)

Poel (1959): BMDS using the erroneous dose adjustment and critical value did identify the BMDL₁₀ that USEPA (2013) has reported, but it is not the model with the best fit of the data. Two BMDL₁₀ values are defined by models that fit the USEPA's input data better. Furthermore, when P>0.1 is used to define an adequate model fit per USEPA guidance, the USEPA-chosen BMDL₁₀ is not defined as meeting them minimum criteria for a model fit.

Poel (1959) Total Tumor Data	USEPA BMDL ₁₀ (µg/day)	Correct BMDL ₁₀ (µg/day)
Lifetime average dose; 100% incidence group removed; P>0.05 to define adequate fit	0.078 (Multistage 3)	0.16*
Lifetime average dose; 100% incidence group removed; P>0.1 to define adequate fit	Not done	0.16*
Lifetime average dose; 100% incidence group included; P>0.1 to define adequate fit	Not done	No fit
Average daily dose; 100% incidence group removed; P>0.1 to define adequate fit	Not done	No fit
Average daily dose; 100% incidence group included; P>0.1 to define adequate fit	Not done	0.216**

* Average of two BMDLs having P>0.1 and AICs of 187 (logistic and probit)

**Average of three BMDLs having P>0.1 and AICs of 186-188 (logistic, probit, quantal-linear).

As noted above, it is entirely erroneous to average the daily doses given to the animals over a longer time period than the animals were alive. This unusual practice has no basis in science or logic, but the practical implications of this dose averaging is to *pretend* that the animals received lower doses as noted below.

Poel (1959) Dose Comparison

Average Daily Doses While Animals Were Alive (µg/day)	Lifetime Average Daily Dose Extending Past Animals' Deaths (µg/day)
0.000	0.000
0.064	0.054
0.163	0.158
0.321	0.237
1.629	0.798
8.143	0.617

There is also no reason to remove the 100% incidence group unless it improves the model fit. As noted above, when the proper dose metric is used for dose-response modeling, the inclusion of the 100% incidence group *improves* the fit.

As is typical, USEPA (2013) has merged benign and malignant tumors. However, the malignant tumor incidence was lower, and the commenters performed benchmark dose modeling to determine the BMDL for malignant tumors.

Poel (1959) Malignant Tumor Data	USEPA BMDL ₁₀ (µg/day)	Correct BMDL ₁₀ (µg/day)
Average daily dose; 100% incidence group removed; P>0.1 to define adequate fit	Not done	0.27 (gamma)
Average daily dose; 100% incidence group included; P>0.1 to define adequate fit	Not done	0.27 (LogLogistic)

In conclusion, USEPA (2013) has modeled the data of Poel (1959) and reported a BMDL₁₀ of 0.078 µg/day. This BMDL₁₀ is in error. The proper BMDL₁₀ for total tumor incidence is 0.216 µg/day. For malignant tumors, the BMDL₁₀ from Poel (1959) is 0.27 µg/day.

Poel (1960) [actually Poel (1963)]: The study of Poel (1960) was investigated. First, the study is mis-cited as Poel (1960) and published in IARC Monograph 10. In fact, the article is Poel (1963) and it was published in *National Cancer Institute Monograph No. 10*. USEPA (2013) reports the data of this paper in Table 1-16 and reports candidate DSFs from the SWR and the C3HeB mice but not the A/He mice in Table 2-11. However, in the Supplemental Information document, the following statement is made and no benchmark dose modeling results are presented.

“For the Poel (1960) studies, all tumors in the highest three dose groups for each of the three mouse strains had occurred by week 40. While these observations support concern for cancer risk, as noted above such results are relatively uncertain for estimating lifetime cancer risk. In addition, there was no information indicating duration of exposure for the mice without tumors; although exposure was for lifetime, it might have been as short as for the mice with tumors. Overall, these datasets did not provide sufficient information to estimate the extent of exposure associated with the observed tumor incidence. Consequently, the experiments reported by Poel (1960) were not used for dose-response modeling.”

USEPA (2013) is confusing when it comes to Poel (1960) [actually Poel (1963)]. As noted above, the Supplemental Document does not list any results, but Table 2-11 of the main document lists a BMDL₁₀ of 0.11 µg/day for both SWR and C3HeB mice.

For SWR mice, the BMDS provided seven model fits with P>0.1, and the best fit is a BMDL₁₀ of 0.13 µg/day from the LogProbit model. USEPA’s 0.11 µg/day (Multistage 3) is associated with a less robust fitting model.

For C3H3B mice, the BMDS provided four model fits with P>0.1 with identical fits (P=0.69, AIC=63) and all yield a BMDL₁₀ of 0.11 µg/day. USEPA’s 0.11 µg/day (Multistage 1) is one the results of one of these four models.

For A/He mice, USEPA (2013) failed to report the model results altogether. BMDS provided nine model fits with P>0.1 with almost identical fits (P=0.88 or 1.0, AIC=16, 18 or 19). The lowest BMDL₁₀ was 1.8 µg/day. The average was 1.96 µg/day.

Poel (1960) [actually Poel (1963)] Total Tumor Data	USEPA BMDL ₁₀ (µg/day)	Correct BMDL ₁₀ (µg/day)
---	--------------------------------------	--

Poel (1960) [actually Poel (1963)] Total Tumor Data	USEPA BMDL ₁₀ (µg/day)	Correct BMDL ₁₀ (µg/day)
SWR- Average daily dose; P>0.05 to define adequate fit	0.11	0.13 (LogProbit)
SWR - Average daily dose; P>0.1 to define adequate fit	Not done	0.13 (LogProbit)
C3HeB - Average daily dose; P>0.05 to define adequate fit	0.11	0.11(Gamma, Multistage 1, Weibull, Quantal-Linear)
C3HeB - Average daily dose; P>0.1 to define adequate fit	Not done	0.11(Gamma, Multistage 1, Weibull, Quantal-Linear)
A/He- Average daily dose; P>0.05 to define adequate fit	Not done	1.96*
A/He- Average daily dose; P>0.1 to define adequate fit	Not done	1.96*

* Average of gamma, logistic, LogLogistic, LogProbit, Multistage Cancer 2, Multistage Cancer 3, Multistage Cancer 4, Probit, and Weibull.

Roe et al. (1970): USEPA (2013) modeled the average daily dose data of Roe et al. (1970) using the very liberal alpha value of 0.05 as the criterion for an adequate fit. By failing to follow USEPA's guidance to use an alpha value of 0.1, the benchmark dose software finds that all 11 models meet the fit criterion, and USEPA chooses the lowest BMDL of 0.39 µg/day as the BMDL₁₀ of choice in Table 2-11. However, this is the BMDL with the lowest P value and the highest AIC value, making it the *worst* BMDL₁₀. ALL other values are from models with better fits to the data. If P>0.05 is used as the cut-off value, the best fit comes from the LogProbit model, with a BMDL value of 0.73 µg/day. In Table E-23, USEPA (2013) does not use the average daily dose, which is the appropriate dose metric. Instead, they use the unjustifiable lifetime averaging dose where they assume that dead animals are alive for 104 weeks. With this inappropriate dose metric and the inappropriate P>0.05 cut-off value, the benchmark dose model gives eleven model fits. The BMDL₁₀ of 0.52 µg/day from the LogProbit model is the best BMDL because the model fit has a higher P value and a lower AIC value than the Multistage 2 model.

Roe et al. (1970) Total Tumor Data	USEPA BMDL ₁₀ (µg/day)	Correct BMDL ₁₀ (µg/day)
Average daily dose; P>0.05 to define adequate fit	0.39 (Table 2-11) (Quantal-Linear)	0.73 (LogProbit)
Average daily dose; P>0.1 to define adequate fit	Not done	0.73 (LogProbit)
Lifetime daily dose; P>0.05 to define adequate fit	0.48 (Table E-23) (Multistage 2)	0.52 (LogProbit)
Lifetime daily dose; P>0.1 to define adequate fit	Not done	0.52 (LogProbit)

USEPA often removes the highest dose group in performing benchmark dose modeling as it did with the data of Poel (1959), but they did not do so here. If the highest dose is removed from the Roe dataset, the fit improves and the BMDL₁₀ from the best fitting model is 0.92 µg/day.

Schmidt et al. (1973): Benchmark dose modeling was performed on the Schmidt et al. (1973) data, and for Swiss mice, the benchmark dose software provides nine models with $P > 0.1$ with AICs of 151, 153, or 155. The best fitting model is the LogProbit model with a $BMDL_{10}$ of 0.22 $\mu\text{g}/\text{day}$. This is the same $BMDL_{10}$ as the one reported by USEPA (2013) from the Multistage 3 model. For NMRI mice, the BMDS gives no best fit model, but nine models had adequate fit, and the average $BMDL_{10}$ was 0.33 $\mu\text{g}/\text{day}$. USEPA's reported $BMDL_{10}$ from the Multistage 2 model is the poorest of the fitting models with a lower P value and a higher AIC.

Schmidt et al. (1973) Total Tumor Data	USEPA $BMDL_{10}$ ($\mu\text{g}/\text{day}$)	Correct $BMDL_{10}$ ($\mu\text{g}/\text{day}$)
Swiss – Average daily dose; $P > 0.05$ to define adequate fit	0.22 (Multistage 3)	0.22 (LogProbit)
Swiss – Average daily dose; $P > 0.1$ to define adequate fit	Not done	0.22 (LogProbit)
NMRI – Average daily dose; $P > 0.05$ to define adequate fit	0.29 (Multistage 2)	0.33*
NMRI – Average daily dose; $P > 0.1$ to define adequate fit	Not done	0.33*

* Average of Gamma, Logistic, LogLogistic, LogProbit, Multistage Cancer 2, 3, and 4, Probit and Weibull.

Schmahl et al. (1977): Benchmark dose modeling was performed on the Schmahl et al. (1977) data, and the benchmark dose software provides 11 model fits with the erroneous $P > 0.05$ cut-off and 10 model fits with the proper $P > 0.1$ cut-off. USEPA (2013) chose the lowest (most potent) $BMDL_{10}$ despite the fact that it has the highest (worst) AIC value. Clearly, the model with the best fit is the LogProbit model with the highest P value and the lowest AIC value, per USEPA guidance.

Schmahl et al. (1977) Total Tumor Data	USEPA $BMDL_{10}$ ($\mu\text{g}/\text{day}$)	Correct $BMDL_{10}$ ($\mu\text{g}/\text{day}$)
Average daily dose; $P > 0.05$ to define adequate fit	0.15 (Multistage 2)	0.24 (LogProbit)
Average daily dose; $P > 0.1$ to define adequate fit	Not done	0.24 (LogProbit)

Habs et al. (1980): Benchmark dose modeling was performed on the Habs et al. (1980) lifetime average data, and the benchmark dose software provides nine models that adequately fit the data using the $P > 0.1$ cut-off criterion. There is no best fit model, but the models with the highest P values, lowest AIC values, and lowest scaled residual include the LogLogistic, LogProbit, and the Probit. The average $BMDL_{10}$ is 0.24 $\mu\text{g}/\text{day}$. USEPA (2013) reports 0.24 $\mu\text{g}/\text{day}$ as the result of the Multistage 4 model, but one cannot run a 4th degree polynomial on a dataset with 4 data points. Table E-23 reports 0.22 $\mu\text{g}/\text{day}$ from a Multistage 3 modeling but BDMS does not give this result.

Benchmark dose modeling was performed on the Habs et al. (1980) average daily data, and the benchmark dose software provides no model fits using the $P > 0.1$ cut-off criterion using all the data. At $P > 0.05$, there are six model fits. The model with the highest P and lowest AIC is the LogProbit model, with a $BMDL$ of 0.18 $\mu\text{g}/\text{day}$. With the highest dose group omitted, BDMS gives eight model fits with $P > 0.1$. The average $BMDL_{10}$ from the four models with $P = 1$ and $AIC = 85$, is 24 $\mu\text{g}/\text{day}$.

Habs et al. (1980) Total Tumor Data	USEPA $BMDL_{10}$ ($\mu\text{g}/\text{day}$)	Correct $BMDL_{10}$ ($\mu\text{g}/\text{day}$)
Lifetime daily dose; $P > 0.05$ to define adequate fit	0.24 (Table 2-11) (Multistage 4) 0.215 (Table E-23)	0.24 (average of LogLogistic, LogProbit and Probit)

	(Multistage 3)	
Lifetime daily dose; P>0.1 to define adequate fit	Not done	0.24 (average of LogLogistic, LogProbit and Probit)
Average daily dose; P>0.05 to define adequate fit	Not done	0.24*
Average daily dose; P>0.1 to define adequate fit	Not done	0.24*

* Average of Gamma, LogLogistic, LogProbit, and Weibull.

Habs et al. (1984): Benchmark dose modeling was performed on the Habs et al. (1984) data, and the benchmark dose software provides 11 model fits using either P>0.05 or P>0.1 as the cut-off criterion. USEPA (2013) has chosen 0.055 µg/day as the BMDL₁₀ citing the Multistage 1 model. The software's Multistage 2 model gives a virtually identical BMDL₁₀ of 0.06 µg/day but this is only one of six models with identical fits, with P=1 and AIC = 48. The average of these six BMDL₁₀ values is 0.068 µg/day.

Habs et al. (1984) Total Tumor Data	USEPA BMDL ₁₀ (µg/day)	Correct BMDL ₁₀ (µg/day)
Average daily dose; P>0.05 to define adequate fit	0.056	0.068*
Average daily dose; P>0.1 to define adequate fit	Not done	0.068*

* Average of six BMDLs with P=1 and AIC = 48 (gamma, LogLogistic, LogProbit, Multistage 2, Multistage Cancer 2, and Weibull)

Grimmer et al. (1983): Benchmark dose modeling was performed on the Grimmer et al. (1983) data, and the benchmark dose software provides seven model fits. All models give P value of 0.92 to 1.0, AIC values of 225 or 227 and scaled residual values of 0. It is appropriate to average the results of all fitting models. The BMDL₁₀ is 0.25 µg/day. USEPA (2013) reports the BMDL₁₀ of 0.21 µg/day from the Multistage 1 model, but this model does not fit the data better than other models.

Grimmer et al. (1983) Total Tumor Data	USEPA BMDL ₁₀ (µg/day)	Correct BMDL ₁₀ (µg/day)
Average daily dose; P>0.05 to define adequate fit	0.21 (Multistage 1)	0.25*
Average daily dose; P>0.1 to define adequate fit	Not done	0.25*

* Average of seven BMDLs with P=0.88 to 1 and AIC = 225 or 227 (gamma, LogLogistic, LogProbit, Multistage Cancer 1,2, Weibull and Quantal-Linear).

Grimmer et al. (1983) Malignant Tumor Data	USEPA BMDL ₁₀ (µg/day)	Correct BMDL ₁₀ (µg/day)
Average daily dose; P>0.05 to define adequate fit	Not done	0.40
Average daily dose; P>0.1 to define adequate fit	Not done	0.40

* Average of four BMDLs with P=0.94 to 0.98 and AIC = 223 (gamma, LogLogistic, LogProbit, Weibull and Quantal-Linear)

Grimmer et al. (1984) [actually Grimmer et al. (1985)]: This dataset provides a tumor incidence of 66% in the lowest dose group. Such a dataset is unsuitable for benchmark dose modeling and should be removed from consideration.

BMDL₁₀ Derivation from Studies not Considered Key Studies: USEPA (2013) has failed to consider several papers that are more suitable for dose-response assessment than the studies selected as key studies. These studies are discussed below.

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

Cavalieri et al. (1983): In this study, BaP was purchased from Aldrich, purified, and recrystallized. Female Swiss mice (28 to 30 per dose group) were given 2.2 nmol, 6.6 nmol or 20 nmol of BaP in 20 µL acetone twice a week for 48 weeks. Survival in control animals was 57 weeks on average. Survival in dosed animals was 55, 58, and 57 weeks for the low, medium and high dose groups.

Cavalieri et al. (1983) Total Tumor Data	USEPA BMDL ₁₀ (µg/day)	Correct BMDL ₁₀ (µg/day)
Average daily dose; P>0.1 to define adequate fit	Not done	0.22 (Multistage Cancer 2)

Levin et al. (1977): In this study, BaP was purchased from Sigma Chemical Company. BaP was given to groups of 30 female C57BL/6J mice in solutions of 50 µL of DMSO: acetone (1:3) except for the high dose of BaP, which was given as two doses of 50 µL each, 30 minutes apart. Doses were 0.025, 0.05, and 0.10 micromoles once every two weeks for 60 weeks.

Levin et al. (1977) Total Tumor Data	USEPA BMDL ₁₀ (µg/day)	Correct BMDL ₁₀ (µg/day)
Average daily dose; P>0.1 to define adequate fit	Not done	0.34 (LogLogistic)

Nesnow et al. (1983): BaP was given to groups of 40 male and 40 female SENCAR mice in 200 µL acetone twice a week for 50-52 weeks. Doses were 12.5, 25.2, 50.5, 101 or 202 µg/week.

Nesnow et al. (1983) Total Tumor Data	USEPA BMDL ₁₀ (µg/day)	Correct BMDL ₁₀ (µg/day)
Average daily dose; P>0.1 to define adequate fit	Not done	1.32 males (LogLogistic) 1.54 females (Weibull)

Summary of BDMS Results

The following table summarizes and compares the BMDL₁₀ values cited by USEPA (2013) and the proper BMDL₁₀ values that result from USEPA's most current BMDS, average daily doses, and a critical value for goodness of fit of alpha = 0.1 as specified in USEPA (2000a) guidance.

Summary of USEPA BMDL₁₀ Values from USEPA (2013) and
from *de novo* Benchmark Dose Modeling

Study	USEPA BMDL ₁₀ (µg/day) Table 2-11	USEPA BMDL ₁₀ (µg/day) Table E-23	Actual BMDL ₁₀ (total tumors) (µg/day)
Poel (1959)	0.078*	0.078*	0.216
Sivak et al. (1997)	0.058*	0.058*	0.076
Poel (1960) (SWR)	0.11	Not listed	0.13
Poel (1960) (C2HeB)	0.11	Not listed	0.11
Poel (1960) (A/He)	Not presented	Not listed	1.96
Roe et al. (1970)	0.39	0.48	0.73 0.92 (highest dose removed)
Schmidt et al. (1970) (Swiss)	0.22	0.22	0.22
Schmidt et al. (1970) (NMRI)	0.29	0.29	0.33
Schmahl et al. (1977)	0.15	0.15	0.24
Habs et al. (1980)	0.24 0.44	0.215	0.24
Habs et al. (1984)	0.056 0.37	0.056	0.068
Grimmer et al. (1983)	0.21 1.0	0.21	0.25
Grimmer et al. (1984)**	0.48 Based on MDML of 70%	0.48 Based on MDML of 70%	Data unsuitable for modeling
Cavalieri et al. (1983)	Not done	Not done	0.22
Levin et al. (1977)	Not done	Not done	0.34
Nesnow et al. (1983) Males	Not done	Not done	1.32
Nesnow et al. (1983) Females	Not done	Not done	1.54

* Values averaged for USEPA's proposed DSF

** Actually Grimmer et al. (1985)

Evaluation of USEPA's Point of Departure Selection

USEPA (2013) has presented 15 candidate DSF values and rejects most of them without cause and has chosen the two highest candidate DSFs to propose. Averaging the candidate DSFs from Sivak et al. (1997) and Poel (1959), they propose $0.005 (\mu\text{g/d})^{-1}$ as the DSF for BaP and other potentially carcinogenic PAHs.

It is not reasonable to omit so much of the data considering that the Sivak et al. (1997) and Poel (1959) studies are the two worst possible studies to use for dose-response modeling given that the Maximum Tolerated Dose was exceeded in both of these studies. The table below shows the resulting DSF if a totally unbiased approach was taken, and the DSF was based on the totality of the data presented in USEPA (2013). The geometric mean DSF is $0.002 (\mu\text{g/d})^{-1}$ and the arithmetic mean DSF is $0.003 (\mu\text{g/d})^{-1}$.

Summary of USEPA-Reported PODs and Candidate DSFs

	Mouse Strain	USEPA Selected Model ^a	BMR	BMD (µg/d)	POD _M =BMDL (µg/d)	Unadjusted Candidate Dermal Slope Factors ^b (µg/d) ⁻¹	POD _{HED} (µg/d)	Adjusted Candidate Dermal Slope Factors ^c (µg/d) ⁻¹
Male mice								
Sivak et al. (1997)	C3H/He _J	Multistage 2°	10%	0.11	0.058	1.7	17.3	0.006
Poel (1959) ^{a,d}	C57L	Multistage 3°	10%	0.13	0.078	1.3	23.3	0.004
Poel (1960) ^{a,d}	SWR	Multistage 3°	10%	0.13	0.11	0.91	32.9	0.003
Poel (1960) ^{a,d}	C3HeB	Multistage 1°	10%	0.16	0.11	0.91	32.9	0.003
Female mice								
Roe et al. (1970)	Swiss	Multistage 2°	10%	0.69	0.39	0.25	116.6	0.001
Schmidt et al. (1973)	Swiss	Multistage 3°	10%	0.28	0.22	0.45	65.8	0.002
Schmidt et al. (1973)	NMRI	Multistage 2°	10%	0.33	0.29	0.34	86.7	0.001
Schmähl et al. (1973)	NMRI	Multistage 2°	10%	0.23	0.15	0.67	44.9	0.002
Habs et al. (1980)	NMRI	Multistage 4°	10%	0.36	0.24	0.42	71.8	0.001
			30%	0.49	0.44	0.69	131.6	0.002
(Habs et al., (1984))	NMRI	Multistage 1°	10%	0.078	0.056	1.8	16.7	0.006
			50%	0.51	0.37	1.4	110.7	0.005
Grimmer et al. (1983)	CFLP	Multistage 1°	10%	0.24	0.21	0.48	62.8	0.002
			40%	1.2	1	0.4	299.1	0.001
Grimmer et al. (1984) ^{a,d}	CFLP	Log-logistic	70%	1.07	0.48	1.5	143.6	0.005
Geometric Mean ^e				0.29	0.20	0.74	60.05	0.002
Arithmetic Mean ^e				0.38	0.25	0.89	75.86	0.003

^aSee Appendix E for modeling details.

^bUnadjusted for interspecies differences. Slope factor=R/BMDLR, where R is the BMR expressed as a fraction.

^cAdjusted for interspecies differences. Cross-species adjustment of dermal doses is based on allometric scaling using the $\frac{3}{4}$ power of body weight. $POD_{HED} (\mu\text{g}/\text{day}) = POD_M (\mu\text{g}/\text{day}) \times (BW_H / BW_M)^{3/4}$.

^dHigh exposure groups omitted prior to dose-response modeling.

^eMean of candidate dermal slope factors, male and female combined. Where two results were provided for a given study (based on two BMR values), these results were averaged before being incorporated into the overall mean.

As discussed elsewhere, USEPA (2013) should reject the Sivak et al. (1997) and Poel (1959) studies as well as the Grimmer et al (1984) study. If these studies are rejected because of exceedance of the Maximum Tolerated Dose and, in the case of Grimmer et al. (1984), because the lowest dose displayed a 66% total cancer incidence, the geometric mean DSF is 0.002 (µg/d)⁻¹ and the arithmetic mean DSF is 0.002 (µg/d)⁻¹.

Summary of USEPA-Reported PODs and Candidate DSFs Excluding Three
Studies

Reference	Mouse Strain	Selected Model ^a	BMR	BMD (µg/d)	POD _M =BMDL (µg/d)	Unadjusted Candidate Dermal Slope Factors ^b (µg/d) ⁻¹	POD _{HED} (µg/d)	Adjusted Candidate Dermal Slope Factors ^c (µg/d) ⁻¹
Male mice								
Poel (1960) ^{a,d}	SWR	Multistage 3°	10%	0.13	0.11	0.91	32.9	0.003
Poel (1960) ^{a,d}	C3HeB	Multistage 1°	10%	0.16	0.11	0.91	32.9	0.003
Female mice								
Roe et al. (1970)	Swiss	Multistage 2°	10%	0.69	0.39	0.25	116.6	0.001
Schmidt et al. (1973)	Swiss	Multistage 3°	10%	0.28	0.22	0.45	65.8	0.002
Schmidt et al. (1973)	NMRI	Multistage 2°	10%	0.33	0.29	0.34	86.7	0.001
Schmähl et al. (1973)	NMRI	Multistage 2°	10%	0.23	0.15	0.67	44.9	0.002
Habs et al. (1980)	NMRI	Multistage 4°	10%	0.36	0.24	0.42	71.8	0.001
			30%	0.49	0.44	0.69	131.6	0.002
(Habs et al., 1984)	NMRI	Multistage 1°	10%	0.078	0.056	1.8	16.7	0.006
			50%	0.51	0.37	1.4	110.7	0.005
Grimmer et al. (1983)	CFLP	Multistage 1°	10%	0.24	0.21	0.48	62.8	0.002
			40%	1.2	1	0.4	299.1	0.001
Geometric Mean ^e				0.31	0.23	0.59	69.51	0.002
Arithmetic Mean ^e				0.36	0.27	0.68	80.68	0.002

^aSee Appendix E for modeling details.

^bUnadjusted for interspecies differences. Slope factor=R/BMDLR, where R is the BMR expressed as a fraction.

^cAdjusted for interspecies differences. Cross-species adjustment of dermal doses is based on allometric scaling using the $\frac{3}{4}$ power of body weight. $POD_{HED} (\mu\text{g/day}) = POD_M (\mu\text{g/day}) \times (BW_H / BW_M)^{3/4}$.

^dHigh exposure groups omitted prior to dose-response modeling.

^eMean of candidate dermal slope factors, male and female combined. Where two results were provided for a given study (based on two BMR values), these results were averaged before being incorporated into the overall mean.

In the following table, the candidate DSFs are presented with (a) the data from Poel (1960) with A/He mice included; (b) the data from Cavalieri et al. (1983), Levin et al. (1977) and Nesnow et al. (1983) included; and (c) the DSFs that result from benchmark dose modeling using the appropriate alpha value for evaluating the goodness of fit and using the BMDL₁₀ from the model with the best fit or, in some cases, the average BMDL₁₀ from multiple models with best fits. Note that Nesnow et al. (1983) is a study performed by USEPA itself. The geometric mean DSF is 0.001 (µg/d)⁻¹ and the arithmetic mean DSF is 0.001 (µg/d)⁻¹.

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.

**Summary of Appropriate PODS and Candidate DSFS from USEPA's Benchmark
Dose Modeling Software**

Reference	Mouse Strain	Selected Model ^a	BM R	POD _M = BMDL (µg/d)	Unadjusted Candidate Dermal Slope Factors ^b (µg/d) ⁻¹	POD _{HED} (µg/d)	Adjusted Candidate Dermal Slope Factors ^c (µg/d) ⁻¹
Male mice							
Poel (1960) ^{a,d}	SWR	LogProbit	10%	0.13	0.77	38.9	0.003
Poel (1960) ^{a,d}	C3HeB	Multiple Fits	10%	0.11	0.91	32.9	0.003
Poel (1960) ^{a,d}	A/He	Multiple Fits	10%	1.96	0.05	586.2	0.0002
Nesnow et al. (1983)	SENCAR	LogLogistic	10%	1.32	0.08	394.8	0.0003
Female mice							
Roe et al. (1970)	Swiss	LogProbit	10%	0.92	0.11	275.1	0.0004
Schmidt et al. (1973)	Swiss	LogProbit	10%	0.22	0.45	65.8	0.002
Schmidt et al. (1973)	NMRI	Multiple Fits	10%	0.33	0.30	98.7	0.001
Schmähl et al. (1973)	NMRI	LogProbit	10%	0.24	0.42	71.8	0.001
Habs et al. (1980)	NMRI	Multiple Fits	10%	0.24	0.42	71.8	0.001
Habs et al. (1984)	NMRI	Multiple Fits	10%	0.068	1.47	20.3	0.005
Grimmer et al. (1983)	CFLP	Multiple Fits	10%	0.25	0.40	74.8	0.001
Cavalieri et al. (1983)	Swiss	Multistage Cancer 2	10%	0.22	0.45	65.8	0.002
Levin et al. (1977)	C57BL/6J	LogLogistic	10%	0.34	0.29	101.7	0.001
Nesnow et al. (1983)	SENCAR	Weibull	10%	1.54	0.06	460.6	0.0002
Geometric Mean				0.34	0.29	101.76	0.001
Arithmetic Mean				0.56	0.44	168.50	0.001

^aSee Appendix E for modeling details.

^bUnadjusted for interspecies differences. Slope factor=R/BMDLR, where R is the BMR expressed as a fraction.

^cAdjusted for interspecies differences. Cross-species adjustment of dermal doses is based on allometric scaling using the ¾ power of body weight. $POD_{HED} (\mu\text{g/day}) = POD_M (\mu\text{g/day}) \times (BW_H / BW_M)^{3/4}$.

^dHigh exposure groups omitted prior to dose-response modeling.

Literature Validation of USEPA (2013) BMDL₁₀ Values

USEPA (2013) has not presented relevant scientific studies that validate the dose-response modeling performed by USEPA.

Higginbotham et al. (1993): In this study, BaP was studied in mice by repeated dermal application to the skin, and no tumors at all were seen. BaP was purified and recrystallized. Doses of 1, 4, or 8 nmol BaP were administered twice weekly in acetone to 23 or 24 female Swiss mice for 40 weeks and then observed until 48 weeks. No carcinomas or papillomas were observed, but in the same experiment high incidences of skin tumors were seen for 7,12-dimethylbenz[a]anthracene and dibenzo[a,l]pyrene at similar doses. The average daily doses of BaP were 0.07, 0.29, and 0.58 µg/day. The USEPA's reported models for the Poel (1959) and the Sivak (1997) data were used to predict the response that USEPA would predict from the Higginbotham et al. (1993) doses.

Higginbotham et al. (1993) Doses	USEPA Predicted Total Tumor Incidence (Poel, 1959)	USEPA Predicted Total Tumor Incidence (Sivak et al., 1997)	Actual Observed Total Tumor Incidence
0.07 µg/day	7%	4%	0%
0.29 µg/day	27%	52%	0%
0.58 µg/day	60%	95%	0%

As seen above, the USEPA models do not agree with the Higginbotham et al. (1993) results at all. However, when the proper average daily dose and the proper goodness of fit cutoff are used to model the Poel (1959) and the Sivak (1997) data, the agreement is better, although both models still overestimate the actual observed tumor incidence, which was 0%.

Higginbotham et al. (1993) Doses	Actual BMDS Model Predicted Total Tumor Incidence (Poel, 1959)	Actual BDMS Model Predicted Total Tumor Incidence (Sivak et al., 1997)	Actual Observed Total Tumor Incidence
0.07 µg/day	0.5%	4%	0%
0.29 µg/day	7%	39%	0%
0.58 µg/day	21%	66%	0%

Recommendations

In conclusion, it is recommended that USEPA (2013) abandon the derivation of a DSF for the many reasons discussed above.

- Among the three studies considered the most relevant and informative for the DSF, the male mouse (reported by Sivak et al., 1997 and Poel, 1959) was more sensitive than the female mouse (reported by Roe et al., 1970). Therefore, the DSF associated with skin tumors in male mice was calculated by linear

extrapolation from the average of the PODs from the Sivak et al. (1997) and Poel (1959) studies. The resulting DSF was selected as the recommended slope factor for assessing human cancer risk following dermal exposure to benzo[a]pyrene. Please comment on whether this selection is scientifically supported and clearly described. Please identify and provide the rationale for any other studies or endpoints that should be selected to serve as the basis for the DSF.

COMMENT:

USEPA (2013) has focused on the Sivak et al. (1997) and Poel (1959) studies as the “best available studies for dose-response analysis”, but as discussed above, these two studies are the *worst* available studies for dose-response analysis because the dosimetry was not quantitative and because Maximally Tolerated Doses were clearly exceeded by evidence of early treatment-related deaths and chronic dermal irritation. In addition, they dismiss the other available listed studies because they claim information is not available that clearly is available. Lastly, USEPA (2013) erroneously rejects for consideration two studies that are of similar or superior quality to all of the Key Studies (Levin et al. [1977] and Nesnow et al. [1983]).

20. The DSF was adjusted to account for interspecies scaling between mice and humans. This cross-species adjustment was based on allometric scaling using body weight to the 3/4 power. Under 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. However, because there is no established methodology for cross-species extrapolation of dermal toxicity, several alternative approaches were evaluated (see Appendix E of the Toxicological Review). Please comment on whether the selected interspecies scaling approach is scientifically supported and clearly described. Also, please comment on whether the alternative approaches presented are clearly described and whether any of these approaches should be selected as the recommended approach. Please identify and provide the rationale for any alternative approach that should be selected.

COMMENT:

Clearly, some interspecies scaling between mice and humans must be done if a DSF were to be derived using mouse skin tumorigenesis data. The allometric scaling performed by USEPA (2013) does not take into account the fact that mouse skin vastly overestimates carcinogenic risks in human skin. These comments present documentation to support the scientific weight of evidence judgment that a DSF should not be derived from mouse skin studies for several reasons.

- Humans exposed to high levels of BaP and other potentially carcinogenic PAHs do not develop skin cancer.
- Mouse skin is known to be more permeable to chemicals, including BaP, than is human skin or other animal skin.
- Mouse skin is known to be more sensitive to PAH-induced skin tumorigenesis than is human skin.

- PAH-induced mouse skin tumors have a different genetic signature than human skin tumors.
- Dermal dosimetry from mouse skin studies is not amenable to or relevant for dose-response modeling for human health risk assessment.

Each of these issues will be discussed separately below.

Humans Exposed to High Levels of BaP and Other Potentially Carcinogenic PAHs do Not Develop Skin Cancer

USEPA (2013) reports the results of three epidemiology studies of coal tar pharmaceutical users: Roelofzen et al. (2010), Pittelkow et al. (1981) and Maughan et al. (1980). The conclusions of these studies are reported correctly: "...high exposure to coal tar treatments was associated with an increased risk of non-melanoma skin cancer." However, USEPA (2013) does not cite several other studies that similarly conclude that long term, high level application of coal tar to human skin does not increase the risk of skin cancer.

Roelofzen et al. (2010) performed an epidemiological study on a cohort of 13,200 patients with psoriasis and eczema. A total of 8,062 of these patients received coal tar treatments. There was no statistically significant increase in overall cancer, skin cancer, internal cancer, or cancer of specific sites, including hematological, breast, lung, gastrointestinal, bladder and urinary tract, prostate, or female reproductive organs observed in this study.

Pittelkow et al. (1981) performed a 25-year follow-up on 280 patients with psoriasis who received coal tar treatments. There was no increase in skin cancer of the coal tar treated individuals compared to expected cancer incidences. The authors stated: "The results of this study suggest that the incidence of skin cancer is not appreciably increased above the expected incidence for the general population when patients are treated with coal tar ointments."

Maughan et al. (1980) performed a 25-year follow-up study on 426 patients who received coal tar ointments clinically. The incidence of skin cancer was not increased above the expected incidence for unexposed populations. The authors' conclusion was: "Our study provides some assurance that the clinical use of coal tar products has not significantly altered the frequency of neoplasms from the natural course." "Those patients in whom skin cancers developed did not receive tar products any longer while hospitalized than did those without skin cancers; nor were they hospitalized more frequently. They did not receive any more coal tar than did the others, and many had received less."

In addition to these three papers, there are several others that reach the same conclusion that are not cited by USEPA (2013). Other papers that conclude that the use of coal tar pharmaceuticals does not increase the risk of skin cancer include:

- Mackenna (1959)

- Muller and Kierland (1964)
- Perry et al. (1968)
- Epstein (1979)
- Muller et al. (1981)
- Bickers (1981)
- Menter and Cram (1983)
- Alderson and Clarke (1983)
- Muller and Perry (1984)
- Lin and Moses (1985)
- Jones et al. (1985)
- Torinuki and Tagami (1988)
- Lindelof and Sigurgeirsson (1993)
- Bhate et al. (1993)
- Jemec and Østerlind (1994)
- Van Schooten and Godschalk (1996)
- Hannukesela-Svahn et al. (2000)

Hannukesela-Svahn et al. (2000) performed an epidemiology study of 5,687 Finnish patients with psoriasis. Coal tar with ultraviolet light treatment was studied (Goeckerman regimen) and there was no statistically significant increase in squamous cell carcinoma or non-Hodgkin's lymphoma in this study.

Van Schooten and Godschalk (1996) was a retrospective study that concluded from reviewing the entire literature that skin cancer has not been increased in psoriasis patients who have been exposed to therapeutically high doses of coal tar.

Jemec and Østerlind (1994) was a long term follow-up study of 88 patients extensively treated with coal tar. All cancers were studied. The authors concluded: "No overall increased risk of cancer was apparent in this

group compared to that of the general population. These data provide further support of the safety of coal tar in the management of dermatological disease.”

The other papers listed above are summarized in the following table.

**Summary of Epidemiological Studies and Review Articles on the Use of
Coal Tar Containing Pharmaceuticals**

Author	Date	Study Population	Conclusion Of Report	Notes Regarding Coal Tar As A Risk Factor
Mackenna (Mackenna RMB, 1959, Uncomplicated Psoriasis, Br Med J, Dec, 1959, 244-1247)	1959	No specific population studied	Adverse effects are rare	Review article concludes that coal tar does not increase risk of skin cancer
Muller and Kierland (Muller SA, Kierland RR, 1964, Crude Coal Tar in Dermatologic Therapy, Mayo Clin Proc, 39, 275-280.)	1964	123 patients treated with coal tar + UV for 38 years	No evidence of adverse effects	Concludes that coal tar is efficacious and safe
Perry et al. (Perry HO, Soderstrom CW, Schulze RW, 1968, The Goeckerman Treatment of Psoriasis, Arch Dermatol, 98, 178-182)	1968	123 patients patients treated with coal tar + UV for 38 years	No evidence of adverse effects	Concludes that coal tar is efficacious and safe
Epstein (Epstein JH, 1979, Risks and Benefits of the Treatment of Psoriasis, New England J Med, 300(15), 852-853)	1979	No specific population studied	Adverse effects are rare despite 50 years of use	Review article concludes that coal tar is not a risk factor for skin cancer.
Maughan et al. (Maughan WZ, Muller, SA, Perry HO, Pittelkow MR and O'Brien PC, 1980, Incidence of Skin Cancers in Patients with Atopic Dermatitis Treated with Coal Tar, Am Acad Dermatol, 3(6), 612-615)	1980	426 patients with atopic dermatitis who received coal tar/UV therapy	Skin cancer was not increased above expected incidence for unexposed populations after 25 years	Concludes that use of coal tar did not increase risk of skin cancer

Author	Date	Study Population	Conclusion Of Report	Notes Regarding Coal Tar As A Risk Factor
Pittelkow et al. (Pittelkow MR, Perry HO, Muller SA, Maughan WZ and O'Brien PC, 1981, Skin Cancer in Patients With Psoriasis Treated With Coal Tar, Arch Dermatol, 117, 465-468.)	1981	280 psoriasis patients who received coal tar/UV therapy	Skin cancer was not increased above expected incidence.	Concludes that use of coal tar did not increase risk of skin cancer
Muller et al. (Muller SA, Perry HO, Pittelkow MR, Maughan WZ, O'Brien PC, 1981, Coal Tar, ultraviolet Light, and Cancer, J Am Acad Dermatol, 4(2), 234-235.)	1981	Patients receiving coal tar/UV treatment	Skin cancer was not increased above expected incidence for unexposed populations	No increase in skin cancer; same patients as in Maughan et al., 1980 and Pittelkow et al., 1981
Bickers (Bickers DR, 1981, The Carcinogenicity and Mutagenicity of Therapeutic Coal Tar - A Perspective, J Invest Dermatol, 77, 173-174)	1981	No specific population discussed	Review article concludes that coal tar is not a risk factor for skin cancer	Review article concludes that coal tar is not a risk factor for skin cancer
Menter and Cram (Menter A and Cram DL, 1983, The Goeckerman Regimen in Two Psoriasis Day Care Centers, J Am Acad Dermatol, 9, 59-65.)	1983	300 psoriasis patients receiving coal tar/UV treatment	No increase in skin cancer compared to expected rates in general population	No increase in skin cancer, although follow-up was short.
Alderson and Clarke (Alderson MR, and Clarke JA, 1983, Cancer Incidence in Patients with Psoriasis, Br J Cancer, 47, 857-859.)	1983	8,405 psoriasis patients with no specific information on treatments	No increase in skin cancer compared to expected rates in general population	Supports conclusion that coal tar does not increase risk of skin cancer, because many patients can be presumed to have received coal tar treatment
Muller and Perry (Muller, S.A. and Perry, H.O. 1984. The Goeckerman Treatment in Psoriasis: Six Decades of Experience at Mayo Clinic. Cutis. 34. 265-269.)	1984	280 psoriasis patients who received coal tar/UV therapy	Skin cancer was not increased above expected incidence.	Concludes that use of coal tar did not increase risk of skin cancer
Lin and Moses (Lin AN, Moses K, 1985, Tar Revisited, Int J Dermatol, 24, 216-218.)	1985	135,000 psoriasis patients	In a survey of 90 dermatologists, only 3 reported skin cancer cases in psoriasis patients	Supports conclusion that coal tar does not increase risk of skin cancer

Author	Date	Study Population	Conclusion Of Report	Notes Regarding Coal Tar As A Risk Factor
Jones et al. (Jones SK, Mackie RM, Hole DJ, Gillis CR, 1985, Further Evidence of the Safety of Tar in the Management of Psoriasis, British Journal of Dermatology, 113, 97-101.)	1985	719 psoriasis patients receiving coal tar therapy only (no psoralens, cytotoxic drugs or UV-B)	No increase in skin cancer seen compared to general population	Supports conclusion that coal tar does not increase risk of skin cancer
Torinuki and Tagami (Torinuki W, Tagami H, 1988, Incidence of Skin Cancer in Japanese Psoriatic Patients Treated with Either Methoxsalen Phototherapy, Goeckerman Regimen, or Both Therapies, J Am Acad Dermatol, 18, 1278-1281.)	1988	43 psoriasis patients who received coal tar/UV therapy	No skin cancers reported.	Supports conclusion that coal tar does not increase risk of skin cancer, although numbers are small and follow-up was short
Lindelof and Sigurgeirsson (Lindelof B, Sigurgeirsson B, 1993, PUVA and Cancer: A Case-Control Study, Br J Dermatol, 129, 39-41.)	1993	24 PUVA skin cancer cases and 96 PUVA controls	Evaluated co-carcinogens with PUVA and found coal tar was not a risk factor	Coal tar did not increase risk of skin cancer even though its use was high.
Bhate et al. (Bhate SM, Sharpe GR, Marks JM, Shuster S, Ross WM, 1993, Prevalence of Skin And Other Cancers in Patients With Psoriasis, Clinical And Experimental Dermatology, 18, 401-4.)	1993	2,247 psoriasis patients receiving coal tar, psoralens, arsenic, methotrexate, and other therapies.	No increased risk seen due to coal tar treatment.	Coal tar was not found to increase the risk of skin cancer.
Jemec & Østerlind (Jemec G.B.E. and A. Østerlind. 1994. Cancer in patients treated with coal tar: a long-term follow up study. <i>J Eur Acad Dermatol Venereol</i> 3:153-156.)	1994	88 patients treated extensively with coal tar from 1917-1937.	No increase in total cancers.	Coal tar was not found to increase the risk of cancer.

Author	Date	Study Population	Conclusion Of Report	Notes Regarding Coal Tar As A Risk Factor
Van Schooten and Godschalk (Van Schooten, F, Godschalk, R. 1996. Coal Tar Therapy: Is it Carcinogenic? Drug Safety 15(6):374-377)	1996	No specific population addressed	No clearly increased skin cancer incidences have been reported in psoriasis patients who have been exposed to therapeutically high doses of coal tar.	Supports conclusion that coal tar does not increase risk of skin cancer because many patients can be presumed to have received coal tar treatment.
Hannuksela-Svahn et al. (Hannukesela-Svahn, A., E. Pukkala, E. Läärä, K. Poikolainen, and J. Karvonen. 2000. Psoriasis, its treatment, and cancer in a cohort of Finnish patients. <i>The Journal of Investigative Dermatology</i> 114(3):587-590)	2000	5,687 psoriasis patients receiving coal tar + UV treatment.	No increase in squamous cell carcinoma or non-Hodgkin's lymphoma above expected levels.	Coal tar was not found to increase the risk of skin cancer or non-Hodgkin's lymphoma.
Roelofzen et al. (Roelofzen, J., K. Aben, U. Oldenhof, P. Coenraads, H. Alkemade, P. van de Kerkhof, P. van der Valk, and L. Kiemeney. 2010. No increased risk of cancer after coal tar treatment in patients with psoriasis or eczema. <i>Journal of Investigative Dermatology</i> 130: 953.)	2010	13,200 patients with psoriasis and eczema. 8,062 received coal tar treatments.	No increase in skin cancer or cancer at other sites above expected levels.	Coal tar was not found to increase the risk of cancer.

In addition, USEPA (2013) gives the results of two irrelevant studies: Stern et al. (1998) and Stern et al. (1980). These two studies are irrelevant to the issue of coal tar exposure, because all of the patients in Stern et al. (1998, 1980) received 8-methoxypsoralen photochemotherapy (PUVA) treatment. PUVA, itself, is known to cause an increased risk of nonmelanoma skin cancer (Stern and Lange [1988] and Gupta et al. [1988]). No information is available in these studies on the cancer rates in patients who received only coal tar treatments.

USEPA (2013) has totally missed the point regarding the coal tar pharmaceutical epidemiological literature. These papers must be considered when summarizing the weight of evidence that BaP may cause skin cancer in humans. It is clear from decades of mouse skin painting experiments that repeated dosing of BaP and BaP-containing mixtures that BaP dissolved in organic solvents and repeatedly painted onto the shaved backs of mice can cause the formation of skin tumors on the mouse skin. This fact is not in dispute. However, these observations are not relevant to human health.

The reasons why such observations are likely to have little if any relevance to human health are numerous.

- Organic solvents dry skin and make it more permeable to large molecules like BaP.
- Mouse skin is more permeable than human skin to organic chemicals.
- Complete carcinogenesis with BaP in mice requires repeated dosing over 50 to 100% of the animal's lifetime.
- Mouse skin is more sensitive to BaP carcinogenesis than is human skin.
- Co-administration of multiple PAHs that are known to cause cancer in mouse skin is *inhibitory*.

In fact, humans are not exposed to BaP in organic solvents every day or several times a week, for 30 to 70+ years. However, some people would be potentially at risk of BaP-induced skin cancer if mouse skin is a good predictor of human skin. These would be people who are intentionally treated with coal tar pharmaceutical products that are high in BaP and other PAHs that have been shown to cause skin cancer in mouse skin. According to the National Psoriasis Foundation, 2.2 percent of the American population has psoriasis. This is 7.5 million Americans.

A high percentage of these Americans are treated by their dermatologists with coal tar pharmaceuticals. Coal tar has been a mainstay of psoriasis treatment for more than 100 years. Because this patient population is exposed to high doses of BaP for many years, the population has been well-studied. Unlike many animal carcinogens for which human epidemiological data are unavailable, this class of chemicals has been well studied. Coal tar pharmaceutical users have been studied by trained epidemiologists, and the results of the studies are uniformly negative. That is, no increase in skin or other cancers has been seen despite the fact that some studies have included up to 8,000 patients, and follow-up periods exceed 25 years.

There is not one published study that has reported an increase in cancer risk among coal tar pharmaceutical users. Studies by Stern and colleagues (Stern et al., 1980, 1998) have been cited as studies showing increases in skin cancer risk, but these studies are irrelevant. They are studies of populations receiving primarily psoralen + UV (PUVA) treatments. A few of these individuals also received coal tar treatments, but because of the confounding primary exposure to PUVA, these studies cannot be used to draw any conclusions about the effects, if any, of coal tar treatment.

Mouse Skin is Known to be More Permeable to Chemicals, Including BaP than is Human Skin or Other Animal Skin

BaP permeates mouse skin with greater efficiency than it does human skin, so the mouse skin is more sensitive to BaP-tumorigenesis by virtue of greater absorption.

Potter et al. (1999) stated: "mouse skin is more penetrable compared to human skin..."

Wester and Maibach (1989) noted that rodent skin is more permeable to chemicals than is human skin. The mouse is 139 times more permeable to paraquat and the hairless mouse is 1,500 times more permeable to paraquat. The order of skin permeability was listed as mouse > guinea pig > goat > rabbit > horse > cat > dog > monkey > weanling pig > man. This is due in part to skin thickness. Hairless mouse skin stratum corneum is two times thinner than the human and mouse skin stratum corneum is three times thinner.

OECD (2010) stated: "Data show that rat, mouse and rabbit skin are generally more permeable than human skin..."

Urano et al. (1995) stated: "Since the number of hair follicles/cm² of mouse skin is much greater than that of human skin and the keratin layer (stratum corneum) of human skin is thicker than that of mouse skin, slower absorption of carcinogens from hair follicles and slower penetration through the stratum corneum in human skin as compared to mouse skin must be considered."

Reifenrath et al. (1984) stated: "Previous studies have shown that the skin of densely haired animals (mice, rats, rabbits, guinea-pigs) tended to be highly permeable, while the permeability properties of the skin of the pig, monkey, and dog were more comparable to that of man (Marzulli et al., 1969; McCreesh, 1965; Tregear, 1964)."

Mouse Skin is Known to be More Sensitive to PAH-Induced Skin Tumorigenesis than is Human Skin

Elegant experiments were done by Graem (1986), Urano et al. (1995), Soballe et al. (1996) and Atilasoy et al. (1997) that show that human skin grafted onto the backs of mice is susceptible to ultraviolet light induced tumorigenesis, but not PAH induced tumorigenesis.

Graem (1986): Human skin grafts on NC nude mice were exposed to two topical applications of 1 mg of DMBA in 50 µL of acetone and/or to applications of 10 µg of TPA in 50 µL of acetone. Tumors did not appear in the central portions of any of the grafts, but epidermal tumors were seen in 34.9% of DMBA treated animals at the graft border. These tumors were from mouse tissue, not human tissue by assessing human blood group B-like antigen and histological staining with bisbenzimidazole. •

Urano et al. (1995) investigated chemical carcinogenesis in human skin using human skin xenografts transplanted to CB-17 SCID mice. 20 µL of 200 nmol DMBA or 300 nmol BaP in acetone was topically applied to human skin grafted onto mice once per week for 25 to 27 weeks. Both DMBA and BaP plus UV irradiation and alternate applications of the carcinogens in combination with UV radiation failed to produce tumors. All treatments induced skin papillomas in the host mouse skin adjoining the grafted human skin. Urano et al. (1995) states: "These results indicate that susceptibility of human skin to these carcinogenic stimuli is much lower than that of mouse skin." "DMBA induced papillomas in allogenic CD4 mouse skin xenografts transplanted to SCID mice..., indicating that the failure of DMBA to induce tumors in human skin xenografts is not due to any damage in association with the transplantation."

Soballe et al. (1996) studied the effects of DMBA, DMBA + UVB, DMBA + UVB + phorbol ester, and UVB alone in human skin grafted onto CB17 SCID mice. The grafted skin survives indefinitely with characteristic human architecture and immunological phenotype and is stable for the life span of the host. DMBA treatment alone caused murine carcinomas in 1 of 16 animals (6%) and zero tumors in human skin. UVB treatment alone caused murine carcinomas in 17 of 38 treated animals (45%) and zero tumors in human skin. DMBA plus UVB treatments caused 13 murine carcinomas in 56 animals (23%) and 2 human carcinomas in 56 animals (3.6%). The authors concluded that their studies demonstrate that “commonly used rodent models may significantly overestimate the human carcinogenic potential of tested agents.”

Atillasoy et al. (1997) was studying the chronic effects of UV light on human skin by grafting human skin onto RAG-1 mice. Mice were treated with UV light alone, DMBA alone, or DMBA + UV light. No papillomas or carcinomas were seen in the RAG-1 mouse xenografts at all.

Kurtz et al. (2004): Kurtz and colleagues found that weekly applications of 200 µg DMBA in 200 µL acetone for six weeks or a single dose of 200 µg DMBA followed by twice weekly applications of TPA (10 µg/200 µL acetone) for six weeks caused 100% incidence of papillomas in adjacent mouse skin on human skin xenografts in CD-17 SCID mice. No papillomas were seen in the identically treated human skin on the xenografted SCID mice. The human skin grafts were viable after 10 weeks. According to the authors: “In general, our present findings suggest that commonly used rodent models could overestimate the carcinogenic potential of agents for humans.” The authors also conclude that the human skin xenograft model has value for the study of human carcinogenesis.

These commenters noticed that these papers were missing from USEPA’s bibliography on BaP and brought them to USEPA’s attention. These papers are important because they present another line of evidence that informs the weight of evidence that human skin is resistant to the skin tumorigenesis that is seen with BaP and other PAHs in mouse skin models.

USEPA’s response to provision of these four papers was to entirely ignore Graem (1986), Soballe et al. (1996), and Atillasoy et al. (1997). They included two sentences on Urano et al. (1995), but then dismissed the findings by stating: “However, it is unclear that this human skin xenograft model preserves the physiological and morphological properties of human skin in vivo (Kappes et al., 2004).” Thus, USEPA (2013) has dismissed the entire human skin xenograft literature in one sentence.

USEPA (2013) correctly reports the conclusions of the work by Kappes et al. (2004). They concluded that human skin grafted to SCID mice do not completely preserve physiological or morphological properties after six months. However, one paper does not define the weight of evidence for any specific topic. Furthermore, the studies of Graem (1986) and Atillasoy et al. (1997) did not employ SCID mice.

More recent literature was obtained and evaluated to determine the weight of evidence regarding the utility of human skin grafted onto mice for the study of human skin carcinogenesis and other dermatotoxicity.

Anna et al. (2007): This paper discusses the results of data from human skin xenograft studies without qualification when summarizing the state of the knowledge about the etiology of human skin cancer. The authors performed a literature review of the mechanism of melanoma and concluded that human skin xenografts are a superior model compared to mouse skin models.

Athar and Kopelovich (2011): This paper discusses the results of data from human skin xenograft studies without qualification when summarizing the effects of a specific agent (rapamycin) in the treatment and prevention of nonmelanoma skin cancer.

Das et al. (1986): Das and colleagues used a human skin xenograft model to show that human skin “maintains its major histological features” and “reserves its metabolic capacity.” They state that xenografts are useful model systems.

Balmain and Harris (2000): This paper concludes that mouse skin tumors have a different genetic signature than do human UV-induced tumors.

Hachiya et al. (2009): The authors used a human skin xenograft model with ICR-SCID mice to study UV exposure on human skin. They concluded that their “skin xenograft model recapitulates premature photoaged skin and provides a comprehensive tool with which to assess the deleterious effects of UVB irradiation.” And “our results suggest that human skin xenograft on SCID mouse is a promising photodamaged skin model.” Finally, they conclude: “Our skin model recapitulates premature photoaged skin and provides a comprehensive tool with which to assess the deleterious effects of UVB irradiation.”

Haftck et al. (1981): The authors used a human skin xenograft model with athymic “nude” BALB/C mice to study morphological and immunological characteristics of human skin. They concluded that immunological properties and keratinization patterns were maintained in the grafts. “These data confirm that the method of grafting human skin on to congenitally athymic nude mice provides an excellent human skin model in subjects other than human beings. Grafted human skin shows no degenerative alternations by morphology, as has been established previously.”

Kim et al. (1992): These authors demonstrated that human skin grafts onto SCID mice maintained a specific disease phenotype and is a useful model for further study.

Nomura et al. (1997): These authors used human skin xenografts on SCID mice and were successful in inducing skin cancer and actinic keratosis from long term UVB exposure in human skin. The authors concluded that human skin xenografts are a useful model of human cancer.

Richmond and Su (2008): Richmond and Su summarized the use of mouse xenograft models including mouse skin models for human cancer therapy and concluded that the xenograft models are “excellent for predicting drug response in human tumors.”

Morton and Houghton (2007): These researchers reviewed the literature and concluded that xenograft models have “for the past two decades, constituted the major preclinical screen for the development of novel

cancer therapeutics. Despite limitations, these models have identified clinically efficacious agents, and remain the 'workhorse' of the pharmaceutical industry."

Zaidi et al. (1992): Zaidi and colleagues studied human skin grafted onto female Balb/c (nu/nu) athymic mice. They found that the grafted human skin was capable of metabolizing activity. Specifically, the grafted human skin could metabolize N-nitrosodimethylamine to active metabolites that were able to react with DNA. The authors concluded that xenografts have utility for skin carcinogenesis research, specifically that "systems using normal human tissues provide opportunities to study effects of a variety of agents on human cells maintained and exposed under normal physiological conditions."

Reed and Manning (1973): These authors demonstrated that full thickness grafts of human skin could be maintained on congenitally athymic (nude) mice for their entire lifetime.

In conclusion, it is clear that the scientific weight of evidence is that human skin xenograft systems in which human skin is grafted onto mouse skin in various mouse strains maintain human characteristics and have great value for the study of human skin tumorigenesis.

PAH-induced Mouse Skin Tumors Have a Different Genetic Signature than Human Skin Tumors

Balmain and Harris (2000) have reviewed the literature and summarized the scientific weight of evidence on carcinogenesis in mouse versus human cells in a paper entitled *Carcinogenesis in mouse and human cells: parallels and paradoxes*. This paper is not cited by USEPA (2013). Balmain and Harris (2000) state that mice are more sensitive to carcinogenic agents than humans. Specifically, they state:

"Why should there be such an apparent speeding up of the whole process of carcinogenesis in mice? Although the mutation frequency is thought to be similar in mouse and human cells, cells of rodent origin are in general much easier to transform in culture, either by treatment with exogenous chemicals or by oncogene transfection. It has been speculated that the difference may be due to less efficient DNA repair, poorer control of genetic stability, or altered control of gene expression through processes such as DNA methylation (for review, see ref. 27)."

Balmain and Harris (2000) present compelling evidence that human skin tumors contain evidence of initiating mutations in the p53 tumor suppressor gene. Similar p53 mutations are seen in the mouse skin tumors caused by UV exposure. However, they are not initiating mutations and are, rather, promoting mutations, because, for instance, p53 null mice do not develop skin tumors unless initiated by chemical agents. The authors present evidence that p53 mutations in the mouse skin induce tumor progression, not tumor initiation. The work of Roop (Greenhalgh et al., 1996) is presented by Balmain and Harris (2000). These researchers found that mice with activated *ras* genes normally develop skin papillomas, but if they are crossed with p53 null mice, no papillomas form. Thus, the *ras* mutation is an initiating mutation. Other evidence is presented showing that PAH-initiated mouse skin tumors show mutations induced in the H-*ras* gene and that the H-*ras* mutations are initiating. In conclusion, PAH-induced mouse skin tumors have a different Mode of Action than do human skin tumors.

Balmain and Harris (2000) also presents a hypothesis that BaP may cause p53 mutations in the lung, so different substances can have different targets in different tissues and even in different target cells within a tissue (Brown et al., 1998). However, the literature strongly suggests that a required event in mouse skin tumorigenesis by BaP and other PAHs is an initiating mutation in H-ras. H-ras mutations have not been reported in human skin tumors. Instead, p53 mutations are implicated in the initiation of human skin cancer, which is caused by ultraviolet light.

USEPA (2013) discusses the fact that BaP induced mutations have been seen in K-ras, H-ras, and p53 targets in various tumor types. The only target mentioned in the skin, is H-ras (Chakravati et al., 1995; Wei et al. 1999), which is consistent with the literature summary of Balmain and Harris (2000). USEPA (2013) discusses p53 as a potential target for BaP in the lung, but there is no mention whatsoever of p53 being associated with human skin tumors. In fact, there is no discussion of the analysis genetic defects in human skin tumors whatsoever.

Recommendations

Because the recent human epidemiology shows that humans exposed to BaP and other potentially carcinogenic PAHs in coal tar pharmaceuticals do not get skin cancer, because human-mouse skin xenograft studies show that human skin is not sensitive to PAH-induced skin tumorigenesis as is mouse skin, because mouse skin is more permeable to PAHs than human skin, and because twice or thrice weekly doses to mouse skin forms an ever increasing skin depot of BaP, any dose-response modeling of mouse skin tumorigenesis results requires a very complicated inter-species scaling approach. Such an approach would require complex modeling to take into account the following factors:

- Greater skin permeation of BaP in mouse skin versus human skin
- Lesser DNA repair activity in mouse skin versus human skin
- Greater promotional mechanisms in mouse skin versus human skin
- Greater sensitivity of mouse skin to chemically induced tumorigenesis versus human skin
- Different mode of action of mouse skin tumorigenesis and human skin tumorigenesis
- Complex dosimetry which takes into account the fact that twice and thrice weekly doses to the mouse skin do not clear and instead form an ever increasing skin depot dose

A. Lack of Real World Validation

1. Sunlight is the Generally Recognized Cause of Human Skin Cancer

Elmets and Athar (2011) stated: "The vast majority of [non-melanoma skin cancers] NMSCs are caused by excessive exposure to UV radiation. In contrast to the stabilized or declined incidences of most other

cancers, the rate of NMSCs continues to rise (2,3). The reasons for this rise relate, at least in part, to increased time people have for outdoor recreational activities, more frequent use of artificial light sources by the lay public for cosmetic purposes, and the increasing proportion of aging individuals in the general population. NMSCs are becoming more frequent in younger people as well.”

Balmain and Harris (2000) have concluded that UV light is the major cause of human skin cancers: “There seems little doubt, given the epidemiological and molecular evidence that has now accumulated, that the culprit for skin cancer induction is indeed UV light.”

In addition, Balmain and Harris (2000) explain in detail in their review article that PAH-induced skin tumors in mouse skin have an H-ras mutation signature whereas human skin cancers have a p53 mutation signature, showing that PAHs are not causally related to human skin cancers.

It is clear that most skin cancer in humans is caused by UV exposure. However, the DSF proposed by USEPA (2013) would predict that much skin cancer is caused by BaP and the other potentially carcinogenic PAHs. 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 following estimates using standard USEPA risk assessment assumptions the population risks posed by (a) touching soil, (b) using coal tar pharmaceuticals, and (c) touching food.

2. Humans Exposed to High Levels of BaP and Other Potentially Carcinogenic PAHs do Not Develop Skin Cancer

As noted in Comment 20.1, extensive epidemiological literature is available on a special population of people who use BaP-containing coal tar pharmaceutical products to treat psoriasis, atopic dermatitis and other skin conditions. These individuals have exceedingly high skin doses of BaP and they have been well-studied. No increases in skin cancer are seen.

In an externally peer reviewed risk assessment report, ICF Consulting (ICF, 2000) estimated that the average total lifetime exposure of coal tar to patients in the Pittelkow et al. (1981) study of individuals being treated for skin conditions with coal tar pharmaceuticals was 254 grams of absorbed PAHs from coal tar. The average daily dose over the lifetime is 254 grams / (70 years * 365 days/year) = 9.9 mg coal tar per day. The BaP-TE content of coal tar can be taken from Culp et al. (1998). The BaP-TE for two coal tar samples was 2,696 ppm and 3,965 ppm. The average is 3,331 ppm or 0.003331. The BaP-TE content of the average daily dose of the coal tar pharmaceutical users can be estimated as (9.9 mg coal tar) x (0.003331 BaP-TE/coal tar) = 0.033 mg BaP-TE per day (33 µg BaP-TE per day).

Assuming the USEPA proposed DSF is correct, the estimated lifetime excess risk of cancer is 1.65E-01. In these studies of thousands of coal tar pharmaceutical users, a risk of 2 in 10 would have easily been detected, but no increases in skin cancer were seen. Assuming USEPA's proposed RPFs, the daily dose increases to 330 µg BaP-TE per day. Assuming the USEPA proposed DSF is correct, the estimated lifetime excess lifetime risk of cancer is 0.8. The proposed DSF is not validated by these real world observations.

3. Human Skin Tumors do Not have the Same Mutational Signature as do Mouse Skin Tumors Induced by B(a)P and other PAHs

As noted in Comment 20.4 above, the weight of evidence points to mutations in the H-ras gene as initiating events in mouse skin tumorigenesis caused by PAHs, such as BaP, but no H-ras mutations are seen in human skin cancers. If BaP was a cause of human skin tumors, then the BaP-signature would be seen in those tumors. The proposed DSF is not validated by the real world observations that human skin tumors do not carry a BaP or other PAH mutational signature.

4. The proposed DSF is Over 100 Times More Potent Than the Proposed OSF

The proposed DSF does not pass even a simple test of logic. If the DSF were true, then the risk posed by ingesting BaP would be inconsequential compared to touching it. This simply is not logical.

5. The Proposed DSF Would Predict a Skin Cancer Epidemic Due to the Ubiquitous Presence of BaP and other PAHs in the Environment

Risk assessment calculations using the proposed DSF indicate that high levels of skin cancer caused by potentially carcinogenic PAHs should be seen in the population.

Touching Soil: The average level of BaP-TE in soil in urban areas (Bradley et al., 1994; USGS, 2003; EPRI, 2003, 2004, 2008) is about 3 mg/kg, but levels of BaP-TE at some sites far exceed this value. BaP-TE levels near roads, railways, highways, parking lots, and other places can exceed 3 mg/kg BaP-TE.

USEPA (2010) has proposed changes to the RPFs used to calculate BaP-TE levels. According to Magee et al. (2012), BaP-TE for coal tar, urban soil, and food will increase by a factor of 10 to 20 based on the known PAHs present, and a factor of far greater when and if concentrations for additional PAHs, which are not current analytes, are added to the BaP-TE calculation. Accordingly, it is assumed for this risk assessment calculation that urban soil contains on average 30 mg/kg BaP-TE. The true level when this RPF document is finalized may be higher.

The lifetime excess cancer risk, assuming the DSF is correct, is $2.1E-02$ for the general population assumed to be exposed to 30 mg/kg BaP-TE. According to the United States Census, about 80% of the United States Population of 350 million lives in urban areas, so the exposed population is 280 million. It is assumed that the remaining 20% of the population living in non-urban areas is exposed to soil containing BaP-TE at levels of 5 mg/kg. Their excess lifetime cancer risk would be $3.6E-03$.

The pro-rated excess lifetime cancer risk for touching soil in the United States for the urban and non-urban population is $1.8E-02$. The lifetime risk of contracting skin cancer is 0.2. So, if USEPA's proposed DSF is true, then BaP in soil is the cause of 8% of all skin cancers. This is not the prevailing view of dermatologists, who have concluded that exposure to ultraviolet light is the cause of almost all skin cancers.

Touching Coal Tar Pharmaceuticals: In an externally peer reviewed risk assessment report, ICF (2000) estimated that the average total lifetime exposure of coal tar to patients in the Pittelkow et al. (1981) study of individuals being treated for skin conditions with coal tar pharmaceuticals was 254 grams of absorbed PAHs from coal tar. The average daily dose over the lifetime is 254 grams / (70 years * 365 days/year) = 9.9 mg coal tar per day. The BaP-TE content of coal tar can be taken from Culp et al. (1998). The BaP-TE for two coal tar samples was 2,696 ppm and 3,965 ppm. The average is 3,331 ppm or 0.003331. The BaP-TE content of the average daily dose of the coal tar pharmaceutical users can be estimated as (9.9 mg coal tar) x (0.003331 BaP-TE/coal tar) = 0.033 mg BaP-TE per day (33 µg BaP-TE per day). Assuming USEPA's proposed RPFs, this value increases to 330 µg BaP-TE per day. Assuming the USEPA proposed DSF is correct, the estimated lifetime excess lifetime risk of cancer is 0.8.

The National Psoriasis Foundation states that 2.2% of the population of the United States has psoriasis. Add atopic dermatitis/other disorders and the fraction would be higher. Assuming 2.2% have psoriasis or other skin conditions, the population that may use coal tar pharmaceuticals is 7.7 million people. If one assumes that 10% of them use coal tar pharmaceuticals, the exposed population is 0.77 million people.

Touching Coal Tar Based Shampoos: ICF (2000) also derived a dose of 5 µg of coal tar absorbed per day from coal tar shampoo use. Assuming the average BaP-TE content of coal tar from above, 3,331 ppm, the dose of BaP-TE from coal tar shampoo use can be estimated as 5 µg/day x 0.003331 = 0.0167 µg/day. Assuming USEPA's proposed RPFs, this value increases to 0.167 µg BaP-TE per day. If the proposed DSF were true, the estimated excess lifetime cancer risk would be 8.35E-04.

Nielsen (1999, as cited in ICF 2000) polled 40,000 households and found that 3,180 had purchased coal tar shampoo (7.95% of people surveyed.) Assuming that 7.95% of the population uses coal tar shampoos, the exposed population is 27 million people.

Estimation of Skin Cancer Cases: In this simple screening population risk assessment, it can be easily estimated that the DSF, if it were a true predictor of human health risk, would predict thousands of cases of cancer just from several of many hundreds of dermal exposures that people have to BaP and other potentially carcinogenic PAHs. Estimates of skin cancer are presented in the table below.

Summary of Estimates of Skin Cancer Cases from Selected Exposures to Potentially Carcinogenic PAHS

Dermal Exposure to BaP	Excess Lifetime Cancer Risk	Number of People Exposed	Number of Lifetime Cases	Lifetime Risk*
Contact with urban soil	2.14E-02 ¹	280 million	5,992,000	
Contact with non-urban soil	3.57E-03 ¹	70 million	249,667	
Use of coal tar pharmaceuticals	8.0E-01 ²	0.77 million	623,700	
Use of coal tar shampoos	8.4E-04 ³	27 million	22,545	
			6,887,912	0.02

*Number of lifetime cases/350 million population

¹ Assumes exposures from USEPA Regional Screening Level calculations (USEPA, 2012): surface area of 5700 cm² for adult and 2800 cm² for child, dermal absorption factor of 0.13, dermal adherence of 0.1 mg/cm² and 0.2 mg/cm² for adult and child, and assuming 70 years exposure. BaP-TE soil concentration assumed to be 3 mg/kg x 10 to account for increased RPFs per USEPA (2010).

² Assumes lifetime average absorbed dose of coal tar is 9.9 mg/day per ICF (2000). BaP-TE concentration assumed to be 3,331 ppm per Culp et al. (1998) x 10 to account for increased RPFs per USEPA (2010).

³ Assumes lifetime average absorbed dose of coal tar is 0.0059 mg/day per ICF (2000). BaP-TE concentration assumed to be 3,331 ppm per Culp et al. (1998) x 10 to account for increased RPFs per USEPA (2010).

Based on the 2013 population of the United States, the excess lifetime skin cancer risk caused by dermal contact with potentially carcinogenic PAHs is estimated to be *at least* 0.02. The lifetime risk of contracting skin cancer is 0.2.

Assuming that USEPA's proposed DSF is correct, the above three common dermal exposures (soil, pharmaceuticals, and shampoo) leads to an estimated lifetime cancer risk of 0.02. The only conclusion that can be drawn from this exercise is that USEPA (2013) believes that potentially carcinogenic PAHs causes 10% of all skin cancer in the United States. If all dermal exposures to all worker groups and the general population were quantitated and summed, it is likely that the proportion of human skin cancer attributed to potentially carcinogenic PAHs *assuming USEPA's proposed DSF were correct*, would approach 100%. This estimate based on EPA's proposed DSF ignores that fact that dermatologists know what causes most skin cancer, and that is exposure to ultraviolet light.

6. The Estimated RSL is Lower than the Levels of BaP and BaP-Toxic Equivalents in Food

The estimated RSL in soil assuming the USEPA (2013) proposed toxicity values is 3.5 µg/kg (ppb). This level of BaP-Toxic Equivalents is similar to or less than the level of BaP-TE in common foods. The European Food Safety Authority reported the levels of many PAHs in foods in EFSA (2008). The upper bound on the mean level of BaP-TE in food throughout Europe is 1.3 to 9.8 µg/kg assuming current or proposed (USEPA, 2010) Relative Potency Factors.

Forsberg et al. (2012) measured PAHs in smoked fish smoked different ways and found that the BaP-TE ranged from 9-27 µg/kg using the current RPFs and 21-64 µg/kg using the proposed RPFs. Ova and Onaran (1998) also measured PAHs in smoked fish and found that BaP-TE concentrations ranged from 2.4 to 4.6 µg/kg in fish and 3.90 to 10.2 µg/kg in eels using current and proposed RPFs respectively.

Djinovic et al. (2008) measured PAHs in smoked meats of various types. BaP-TE concentrations ranged from 1.6 to 217 µg/kg in smoked bacon using current and proposed RPFs respectively. BaP-TE concentrations in other smoked meats were comparable.

Many other papers can be cited that show that the estimated RSLs assuming the proposed toxicity factors would be lower than the BaP-TE levels in many foods.

B. Lack of Discussion of Policy Implications

USEPA (2013) has not thought ahead about the implications of these proposed toxicity values. The most striking implications of the proposal are due to the proposed DSF. If this DSF, which is shown above to be scientifically incorrect on many counts, were released, it would be used in the next update of the Regional Screening Levels (RSLs). The RSLs are widely used throughout the country as screening levels and also as *de facto* clean up levels, despite the fact that the RSL guidance manual states that they should not be used as cleanup levels.

The RSL for BaP, which is used as the indicator for all potentially carcinogenic PAHs with USEPA Relative Potency Factors, is currently 0.02 mg/kg for residential soil and 0.2 mg/kg for industrial/commercial soil. These RSLs are already lower than the background levels of PAHs in many localities. Although cleanup is required in some jurisdictions to a 1×10^{-6} excess lifetime cancer risk level, in many jurisdictions, the target risk level for waste site cleanups is 1×10^{-5} . Even with this target risk level, the *de facto* cleanup levels for potentially carcinogenic PAHs are 0.2 mg/kg for residential land use and 2 mg/kg for industrial/commercial land use as BaP-Toxic Equivalent (BaP-TE) concentrations. These levels are lower than typical urban background levels for BaP-TE.

So, the current toxicity factors and the current risk assessment methods and procedures already dictate that soil cleanups are required for soils that are not higher in potentially carcinogenic PAHs than soils throughout the country in areas not affected by waste sites or waste releases.

The proposed DSF would cause the RSLs to drop to 0.0035 mg/kg or 3.5 parts-per-billion for residential land use and 0.031 mg/kg or 31 parts-per-billion for industrial/ commercial land use. The implications of these values are staggering. All sites in the entire country will be screened *in* for site investigations and risk assessments, and cleanups will be required for all residential and commercial properties in the entire country and perhaps national parks and wilderness areas which are affected at an increasing frequency by forest fires which leave PAHs in the soils.

In addition, all sites regardless of the nature of the released constituents will become “PAH sites” because of the ubiquitous nature of PAHs in the environmental from both natural and anthropogenic sources.

Additionally, all sites that have been remediated, even if they were not “PAH sites” will be re-opened when the Five Year Reviews compare the post-remediation data to the new criteria driven entirely by the new DSF for BaP.

In conclusion, the proposed DSF (2013) BMDL10 Values is scientifically flawed as discussed in previous sections of these comments, and it should be abandoned.

Summary of Proposed Toxicity Factors

Toxicity Factor	Existing	Proposed	Difference
-----------------	----------	----------	------------

Inhalation Unit Risk	None (0.0011 (µg/m ³) ⁻¹ [a])	0.0005 (µg/m ³) ⁻¹	2x less potent
Oral Cancer Slope Factor	7.3 (mg/kg-day) ⁻¹	1.0 (mg/kg-day) ⁻¹	7x less potent
Dermal Cancer Slope Factor	None (7.3 (mg/kg-day) ⁻¹ [b])	350 (mg/kg-day) ⁻¹ [c]	48x more potent
Oral Reference Dose	None	0.0003 mg/kg-day	--
Inhalation Reference Concentration	None	0.000002 mg/m ³	--

[a] There is no IUR in USEPA's IRIS database, but the USEPA Regional Screening Levels use this IUR, which was derived by the State of California. This value is not used by all USEPA programs.

[b] The oral cancer slope factor is used as a surrogate for the dermal pathway.

[c] The Draft Toxicological Review of Benzo(a)pyrene indicates the proposed dermal slope factor is 0.005 (µg/day)⁻¹. DSF listed above assumes 70 kg adult.

Summary of Proposed Regional Screening Levels

Regional Screening Level	Existing	Proposed	Difference
Residential Soil RSL [a]	0.015 mg/kg (ca)	0.0035 mg/kg (ca)	4x lower
Industrial Soil RSL	0.21 mg/kg (ca)	0.031 mg/kg (ca)	7x lower
Residential Ambient Air RSL [a]	0.00087 µg/m ³ (ca)	0.0019 µg/m ³ (ca)*	2x higher
Industrial Ambient Air RSL	0.011 µg/m ³ (ca)	0.0088 µg/m ³ (nc)	2x lower
Tapwater RSL [a]	0.0029 µg/L (ca)	0.022 µg/L (ca)	7x higher
Fish Tissue RSL	0.00043 mg/kg (ca)	0.0032 mg/kg (ca)	7x higher

[a] RSL based on cancer risk includes adjustment for early-life susceptibility.

(ca) = RSL is based on a cancer risk level of 1×10^{-6}

(nc) = RSL is based on a hazard quotient of 1.

* If a risk level of 1×10^{-5} is used to calculate the carcinogenic RSL, the non-cancer RSLs would be lower than the cancer RSLs.

These proposed toxicity values will have wide ranging implications for all USEPA and state environmental regulatory programs, not just CERCLA and RCRA soil remediation programs. All USEPA programs rely on the IRIS database as the basis of their regulations: air, water, waste management, etc. A partial list of the rules and regulations that will be affected by this proposal includes:

- CAA Maximum Achievable Control Technology Standards

- CAA New Source Performance Standards
- CAA National Ambient Air Quality Standards
- CAA Title V Operating Permits
- CAA New Source Review Construction Permit Program
- CAA Prevention of Significant Deterioration Permit Program
- CWA Maximum Contaminant Levels
- RCRA Waste Classification
- SARA Title III Reporting

III. References

Alderson, M.R. and J.A. Clarke. 1983. Cancer incidence in patients with psoriasis. *Br J Cancer* 47:857-859.

Anna, B., B. Zbytek, J. Granese, J.A. Carlson, J. Ross, and A. Slominski. 2007. Mechanism of UV-related carcinogenesis and its contribution to nevi/melanoma. NIH Public Access Author Manuscript Published in final edited form as *Expert Rev Dermatol* 1007; 2(4): 451-469.

API et al. 2010. Comments on Docket ID No. USEPA-HQ-ORD-2010-0047, Development of a Relative Potency Factor (RPF) Approach for Polycyclic Aromatic Hydrocarbon (PAH) Mixtures Submitted by the Following: American Coke and Coal Chemicals Institute, American Petroleum Institute, Asphalt Institute, Association of American Railroads, National Petrochemical & Refiners Association & Pavement Coatings Technology Council. April 27, 2010.

Archibong, A.E., F. Inyang, A. Ramesh, M. Greenwood, T. Nayyar, P. Kopsombut, D.B. Hood, and A.M. Nyanda. 2002. Alteration of pregnancy related hormones and fetal survival in F-344 rats exposed by inhalation to benzo(a)pyrene. *Reprod Toxicol* 16: 801-808.

Archibong, A.E., A. Ramesh, M.S. Niaz, C.M. Brooks, S.I. Roberson, and D.D. Lunstra. 2008. Effects of benzo(a)pyrene on intra-testicular function in F-344 rats. *Int J Environ Res Public Health* 5: 32-40.

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

Athar, M. and L. Kopelovich. 2011. Rapamycin and mTORC1 Inhibition in the Mouse: Skin Cancer Prevention. *Cancer Prevention Research* 4:957-961.

- Atilasoy, E.S., R. Elenitsas, E.R. Sauter, P.W. Soballe, and M. Herly. 1997. UAB Induction of Epithelial Tumors in Human Skin Using a RAG-1 Mouse Xenograft Model. *J. Invest. Dermatol* 109:704-709.
- Baan, R., Y. Grosse, K. Straif, B. Secretan, F. El Ghissassi, V. Bouvard, L. Benbrahim-Tallaa, N. Guha, C. Freeman, L. Galichet, and V. Coglianò. 2009. A review of human carcinogens--part F: Chemical agents and related occupations. *Lancet Oncol* 10: 1143-1144.
- Balmain, A. and C.C. Harris. 2000. Carcinogenesis in mouse and human cells: parallels and paradoxes. *Carcinogenesis* 21:371-377.
- Beland, F., and S. Culp. 1998. Chronic bioassay of two composite samples from selected manufactured gas plant waste sites [unpublished report]. (Technical Report 6722.02). Jefferson, AK: National Center for Toxicological Research.
- Benbrahim-Tallaa, L., R.A. Baan, Y. Grosse, B. Lauby-Secretan, F. El Ghissassi, V. Bouvard, N. Guha, D. Loomis, and K. Straif. 2012. Carcinogenicity of diesel-engine and gasoline-engine exhausts and some nitroarenes. *Lancet Oncol* 13: 663-664.
- Bergdahl I.A. and B. Jarvholm. 2003. Cancer morbidity in Swedish asphalt workers. *Am J Ind Med* 43: 104–108.
- Berger, J. and A. Manz. 1992. Cancer of the stomach and the colon-rectum among workers in a coke gas plant. *Am J Ind Med* 22: 825-834.
- Bhate, S.M., G.R. Sharpe, J.M. Marks, S. Shuster, and W.M. Ross. 1993. Prevalence of skin and other cancers in patients with psoriasis. *Clin Exp Dermatol* 18:401-4.
- Bickers, D.R. 1981. The Carcinogenicity and mutagenicity of therapeutic coal tar – A perspective. *J Invest Dermatol* 77:173-174.
- Bjor, O., L. Damberg, C. Edstrom, and T. Nilsson. 2008. Longterm follow-up study of mortality and the incidence of cancer in a cohort of workers at a primary aluminum smelter in Sweden. *Scand J Work Environ Health* 34:463–470.
- Bradley, L.J.N., B.H. Magee, and S.L. Allen. 1994. Background Levels of Polycyclic Aromatic Hydrocarbons (PAH) and Selected Metals in New England Urban Soils. *Journal of Soil Contamination*.3:1-13.
- Brown, K., D. Strathdee, S. Bryson, W. Lambie and A. Balmain. 1998. The malignant capacity of skin tumors induced by expression of a mutant H-ras transgene depends on the cell type targeted. *Curr. Biol.* 8:516-524.
- Brown, J.R. and J.L. Thornton. 1957. Percivall Pott (1714-1788) and chimney sweepers' cancer of the scrotum. *Br J Ind Med* 14: 68-70.

Buck C. & Reid, DD. 1956. Cancer in coking plant workers. *Br J Ind Med*, 13: 265–269.

Butlin, H.T. 1892. Three Lectures on Cancer of the Scrotum in Chimney-Sweeps and Others: Delivered at the Royal College of Surgeons of England. *British Medical Journal* 2: 66–71.

Carta, P., G. Aru, C. Cadeddu, G. Gigli, G. Papi, F. Carta, and P. Nurchis. 2004 Mortality for [sic] pancreatic cancer among aluminium smelter workers in Sardinia, Italy. *G. Ital. Med. lav. Ergon.* 26, 83–89.

Cavalieri, E., A. Munhall, E. Rogan, S. Salmasi, and K. Patil. 1983. Syncarcinogenic effect of the environmental pollutants cyclopenteno[cd]pyrene and benzo[a]pyrene in mouse skin. *Carcinogenesis* 4, 393–397

Chakravarti, D., J.C. Pelling, E.L. Cavalieri, and E.G. Rogan. 1995. Relating aromatic hydrocarbon-induced DNA adducts and c-H-ras mutations in mouse skin papillomas: The role of apurinic sites. *PNAS* 92: 10422-10426.

Chau, N., J.P. Bertrand, J.M. Mur, A. Figueredo, A. Patris, J.J. Moulin, and Q.T. Pham. 1993. Mortality in retired coke oven plant workers. *Br J Ind Med* 50: 127-135.

Chen, G., J. Gingerich, L. Soper, G.R. Douglas, and P.A. White. 2010. Induction of lacZ mutations in MutaMouse primary hepatocytes. *Environ Mol Mutagen* 51: 330-337.

Chen, X., H. An, L. Ao, L. Sun, W. Liu, Z. Zhou, Y. Wang, and J. Cao. 2011. The combined toxicity of dibutyl phthalate and benzo(a)pyrene on the reproductive system of male Sprague Dawley rats in vivo. *J Hazard Mater* 186: 835-841.

Chen, C., Y. Tang, X. Jiang, Y. Qi, S. Cheng, C. Qiu, B. Peng, and B. Tu. 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.

Costantino, J.P., C.K. Redmond, and A. Bearden. 1995. Occupationally related cancer risk among coke oven workers: 30 years of follow-up. *J Occup Environ Med* 37: 597-604.

Cookson, H.A. 1924. Epithelioma of the skin after prolonged exposure to creosote. *Br. Med. J.*, i,368.

Culp, S.J., D.W. Gaylor, W.G. Sheldon, L.S. Goldstein, and F.A. Beland. 1998. A comparison of the tumors induced by coal tar and benzo[a]pyrene in a 2-year bioassay. *Carcinogenesis* 19: 117-124.

Das, M., P. Asokan, P.S.C. Don, G. Krueger, D.R. Bickers, and H. Muktar. 1986. Carcinogen Metabolism In Human Skin Grafted Onto Athymic Nude Mice: A Model System for the Study of Human Skin Carcinogenesis. *Biochemical and Biophysical Research Communications* 138: 33-39.

Davies, G.M. 1977. A mortality study of coke oven workers in two South Wales integrated steelworks. *Br J Ind Med*. 34: 291–297.

De Jong, W.H., E.D. Kroese, J.G. Vos, and H. Van Loveren. 1999. Detection of immunotoxicity of benzo[a]pyrene in a subacute toxicity study after oral exposure in rats. *Toxicol Sci* 50: 214-220.

Djinovic, Jasna, Aleksandar Popovic, and Wolfgang Jira . 2008. Polycyclic aromatic hydrocarbons (PAHs) in different types of smoked meat products from Serbia. *Meat Science* 80:449–456.

Doll, R., M.P. Vessey, R.W.R. Beasley, A.R. Buckley, E.C. Fear, R.E.W. Fisher, E.J. Gammon, W. Gunn, G.O. Hughes, K. Lee, and B. Norman-Smith. 1972. Mortality of gasworkers — Final report of a prospective study. *Br. J. Ind. Med.* 29, 394–406

Donato, F., S. Monarca, G. Marchionna, A. Rossi, C. Cicioni, R. Chiesa, D. Colin, and P. Boffetta. 2000. Mortality from cancer and chronic respiratory diseases among workers who manufacture carbon electrodes. *Occup. environ. Med.* 57, 484–487

Earle, J. 1808. *The Chirurgical Works of Percivall Pott, F.R.S.*, Vol. III. London: Wood & Innes, pp. 176–183.

European Food Safety Authority (EFSA). 2008. Scientific opinion of the panel on contaminants in the food chain on a request from the European Commission on Polycyclic Aromatic Hydrocarbons in Food. *The EFSA Journal* 724:1-114.

Elmets, C.A. and M. Athar. 2011. Targeting Ornithine Decarboxylase for the Prevention of Nonmelanoma Skin Cancer in Humans Department of Dermatology, UAB Skin Diseases Research Center and UAB Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, Alabama. NIH Public Access Author Manuscript *Cancer Prev Res (Phila)*. 2010 January; 3(1): 8-11.

Epstein J.H. 1979. Risks and Benefits of the treatment of psoriasis. *New England J Med* 300(15):852-853.

Electric Power Research Institute (EPRI). 2003. Polycyclic Aromatic Hydrocarbons (PAHs) in Surface Soils in Western New York. 1005296. October.

Electric Power Research Institute (EPRI). 2004. Polycyclic Aromatic Hydrocarbons (PAHS) In Surface Soil in Illinois. Background PAHs. 1011376. December.

Electric Power Research Institute (EPRI). 2008. Examination of the Sources of Polycyclic Aromatic Hydrocarbon (PAH) in Urban Soil. Electric Power Research Institute. 1015558. December.

Epstein, J.H. 1979. Risks and benefits of the treatment of psoriasis. *New England J Med* 300(15):852-853.

Evanoff, B.A., P. Gustavsson, and C. Hogstedt. 1993. Mortality and incidence of cancer in a cohort of Swedish chimney sweeps: an extended follow up study. *Br J Ind Med*, 50:450–459.

Fasano, W.J. and J.N. McDougal. 2008. In vitro dermal absorption rate testing of certain chemicals of interest to the Occupational Safety and Health Administration: Summary and evaluation of USEPA's mandated testing. *Regulatory Toxicology and Pharmacology* 51:181-194.

FDA (U.S. Food and Drug Administration). 2001. Letter from Dennis E. Baker, Associate Commissioner for Regulatory Affairs, FDA to Perry M. Gottesfeld, Occupational Knowledge International. Docket No. 00P-1210. February.

Finnish Forest Research Institute. 2003. Finnish Statistical Yearbook of Forestry

Forsberg, N.D., D. Stone, A. Harding, B. Harper, S. Harris, M.M. Matzke, A. Cardenas, K.M. Waters, and K.A. Anderson. 2012. Effect of Native American fish smoking methods on dietary exposure to polycyclic aromatic hydrocarbons and possible risks to human health. *Journal of Agricultural and Food Chemistry* 60:6899-6906.

Friesen, M.C., P.A. Demers, and J.J. Spinelli. 2007. Comparison of two indices of exposure to polycyclic aromatic hydrocarbons in a retrospective aluminium smelter cohort. *Occup Environ Med* 64: 273-278.

Friesen, M.C., G. Benke, and A. Del Monaco. 2009. Relationship between cardiopulmonary mortality and cancer risk and quantitative exposure to polycyclic aromatic hydrocarbons, fluorides, and dust in two prebake aluminum smelters. *Cancer Causes Control* 20: 905-916.

Gao, M., Y. Li, Y. Sun, J. Long, Y. Kong, S. Yang, and Y. Wang. 2011. A common carcinogen benzo[a]pyrene causes p53 overexpression in mouse cervix via DNA damage. *Mutation Research* 724:69– 75

Gibbs, G.W., B. Armstrong, and M. Sevigny. 2007. Mortality and cancer experience of Quebec aluminum reduction plant workers. Part 2: mortality of three cohorts hired on or before January 1, 1951. *J Occup Environ Med* 49: 1105-1123.

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

Gibbs, G.W. and M. Sevigny. 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.

Giovanazzi, A. and F. D'Andrea. 1981. Causes of death among workers in an aluminium electrolytic reduction plant. *Med. Lav.* 72:277–282

Graem, N. 1986. Epidermal Changes following Application of 7,12-Dimethylbenz(a)anthracene and 12-O-Tetradecanoylphorbol-13-acetate to Human Skin Transplanted to Nude Mice Studied with Histological Species Markers. *Cancer Res* 46:278-284

Greenhalgh, D.A., Wang, X.J., Donehower, L.A., and Roop, D.R. 1996. Paradoxical tumor inhibitory effect of p53 loss in transgenic mice expressing epidermal-targeted *v-rasHa*, *v-fos*, or human transforming growth factor alpha. *Cancer Res*.56:4413-4423. As cited in Balmain and Harris (2000).

Grimmer, G., H. Brune, R. Deutsch-Wenzel, K.W. Naujack, J. Misfeld, and J. Timm. 1983. On the contribution of polycyclic aromatic hydrocarbons to the carcinogenic impact of automobile exhaust condensate evaluated by local application onto mouse skin. *Cancer Lett* 21: 105-113.

Grimmer, G., H. Brune, R. Deutsch-Wenzel, G. Dettbarn, and J. Misfeld. 1984. Contribution of polycyclic aromatic hydrocarbons to the carcinogenic impact of gasoline engine exhaust condensate evaluated by implantation into the lungs of rats. *J Natl Cancer Inst* 72: 733-739.

Grimmer, G., H. Brune, R. Deutsch-Wenzel, G. Dettbarn, J. Misfeld, U. Abel, and 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

Gupta, A.K., R.S. Stern, N.A. Swanson, and T.F. Anderson. 1988 PUVA follow-up study: Cutaneous melanomas in patients treated with psoralens plus ultraviolet A. *J Am Acad Dermatol* 19: 67-76.

Gustavsson P. and Reuterwall, C. 1990. Mortality and incidence of cancer among Swedish gas workers. *British Journal of Industrial Medicine* 47:169-174.

Gustavsson, P., T. Bellander, L. Johansson, and S. Salmonsson. 1995. Surveillance of mortality and cancer incidence among Swedish graphite electrode workers. *Environ. Res.* 70, 7-10

Habs, M., D. Schmähl, and J. Misfeld. 1980. Local carcinogenicity of some environmentally relevant polycyclic aromatic hydrocarbons after lifelong topical application to mouse skin. *Arch Geschwulstforsch* 50: 266-274.

Habs, M., S.A. Jahn, and D. Schmähl. 1984. Carcinogenic activity of condensate from colocynth seeds (*Citrullus colocynthis*) after chronic epicutaneous administration to mice. *J Cancer Res Clin Oncol* 108: 154-156.

Hachiya, A., P. Sriwiriyanont, T. Fujimura, A. Ohunchi, T. Kitahara, Y. Takema, W.J. Kitzmiller, M.O. Visscher, R. Tsuboi, and R.E. Boissy. 2009. Mechanistic Effects of Long-Term ultraviolet B Irradiation Induce Epidermal and Dermal Changes in Human Skin Xenografts. *The American Journal of Pathology*, 174:401-413.

- Hafttek, M., J-P. Ortonne, M-J. Staquet, J. Viac, and J. Thivolet. 1981. Normal and Psoriatic Human Skin Grafts on "Nude Mice": Morphological and Immunochemical Studies. *Journal of Investigative Dermatology* 76:48-52.
- Haldorsen, T., A. Andersen, and P. Boffetta. 2004 Smoking-adjusted incidence of lung cancer by occupation among Norwegian men. *Cancer Causes Control* 15:139–147
- Hammond, E.C., I.J. Selikoff, P.L. Lawther, and H. Seidman. 1976. Inhalation of benzpyrene and cancer in man. *Ann N Y Acad Sci* 271: 116-124.
- Hannukesela-Svahn, A., E. Pukkala, E. Läärä, K. Poikolainen, and J. Karvonen. 2000. Psoriasis, its treatment, and cancer in a cohort of Finnish patients. *The Journal of Investigative Dermatology* 114(3):587-590.
- Hansen, K.S., S. Viskum, and M.S. Pedersen. 1986. Mortality among gas workers. *Ugeskr. Laeger* 148, 610–612 (in Danish)
- Hansen, E.S. 1989. Cancer incidence in an occupational cohort exposed to bitumen fumes. *Scand J Work Environ Health* 15:101-5.
- Hansen, E.S. 1991. Mortality of mastic asphalt workers. *Scand J Work Environ Health* 17:20-4.
- Henry, S.A. 1937. The study of fatal cases of cancer of the scrotum from 1911 to 1935 in relation to occupation, with special reference to chimney sweeping and cotton mule spinning. *The American Journal of Cancer* 31:28–57.
- Henry, S.A. 1946. *Cancer of the Scrotum in Relation to Occupation*. (Oxford University Press: New York).
- Henry, S.A. 1947. Occupational cutaneous cancer attributable to certain chemicals in industry. *British Medical Bulletin* 4: 389–401.
- Henry, S.A. and E.D. Irvine. 1936. Cancer of the scrotum in the Blackburn registration district, 1837-1929. *The Journal of Hygiene* 36: 310-340.
- Higginbotham, S., N.V.S. Ramakrishna, S.L. Johansson, E.G. Rogan, E.L. Cavalieri. 1993. Tumor-initiating activity and carcinogenicity of dibenzo[a,l]pyrene versus 7,12-dimethylbenz[a]anthracene and benzo[a]pyrene at low doses in mouse skin. *Carcinogenesis* 14: 875-878.
- Hogg, S. 1996. A Review of the Validity and Variability of the Elevated Plus-Maze as an Animal Model of Anxiety. *Pharmacology Biochemistry and Behavior* 54(1): 21-30.
- Hogstedt, C., C.Jansson, M. Hugosson, H.Tinnerberg, and P.Gustavsson. 2013. Cancer Incidence in a Cohort of Swedish Chimney Sweeps, 1958–2006. *American Journal of Public Health*. 103:1708-1714.

Hueper, W.C. 1942. Cancers of the Respiratory System. In Occupational Tumors and Allied Diseases (Charles Thomas; New York). Pages 420-435.

Hurley, J.F., R.M. Archibald, P.L. Collings, D.M. Fanning, M. Jacobsen, and R.C. Steele. 1983. The mortality of coke workers in Britain. *Am. J. ind. Med.* 4, 691–704

International Agency for Research on Cancer (IARC). 1985. Polynuclear aromatic compounds, Part 4, bitumens, coal-tars and derived products, shale-oils and soots. *IARC Monogr Eval Carcinog Risk Chem Hum* 35: 1–247.

IARC. 1987. Overall evaluations of carcinogenicity: an updating of IARC Monographs volumes 1 to 42. *IARC Monogr Eval Carcinog Risks Hum Suppl* 7: 1–440.

IARC. 2010a. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 92. Some Non-heterocyclic Polycyclic Aromatic Hydrocarbons and Some Related Exposures*. World Health Organization (WHO).

IARC. 2010b. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 95. Household Use of Solid Fuels and High-Temperature Frying*. World Health Organization (WHO).

IARC. 2012a. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 100E. personal habits and indoor combustions*. World Health Organization (WHO).

IARC. 2012b. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 100F. Chemical Agents and Related Occupations: A Review of Human Carcinogens*. World Health Organization (WHO).

IARC. 2013. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 103. Bitumens and Bitumen Emissions, and Some N- and S-Heterocyclic Polycyclic Aromatic Hydrocarbons*. World Health Organization (WHO).

ICF Consulting. 2000. *Estimation of Lifetime Skin Cancer Risk from the Use of Coal Tar-Containing Shampoos*. Prepared by The K.S. Crump Group, Inc. and ICF Consulting. 602 East Georgia Street, Ruston, Louisiana 71270. Reviewed by TERA Peer Review Panel. September.

Jemec, G.B.E. and A. Østerlind. 1994. Cancer in patients treated with coal tar: a long-term follow up study. *J Eur Acad Dermatol Venereol* 3:153-156.

Jones S.K., R.M. Mackie, D.J. Hole, and C.R. Gillis. 1985. Further Evidence of the Safety of Tar in the Management of Psoriasis. *British Journal of Dermatology* 113, 97-101.

Jules, G.E., S. Pratap, A. Ramesh, and D.B. Hood. 2012. In utero exposure to benzo(a)pyrene predisposes offspring to cardiovascular dysfunction in later-life. *Toxicology* 295: 56-67.

Kao, J., J. Hall, and G. Helman. 1988. *In Vitro* Percutaneous Absorption in Mouse Skin: Influence of Appendages. *Toxicol Appl Pharmacol* 94:93-104.

Kappes, U., S. Schliemann-Willers, L. Bankova, C. Heinemann, T.W. Fischer, M. Ziemer, H. Schubert, J. Norgauer, J.W. Fluhr, and P. Elsner. 2004. The quality of human skin xenografts on SCID mice: A noninvasive bioengineering approach. *Br J Dermatol* 151: 971-976.

Karlehagen, S., A. Andersen, and C.G. Ohlson. 1992. Cancer incidence among creosote-exposed workers. *Scand J Work Environ Health* 18: 26-29.

Kauppinen, T., P. Heikkilä, T. Partanen, S.V. Virtanen, E. Pukkala, P. Ylöstalo, I. Burstyn, G. Ferro, and P. Boffetta. 2003. Mortality and Cancer Incidence of Workers in Finnish Road Paving Companies. *American Journal of Industrial Medicine* 43:49-57

Kawai, M., H. Amamoto, and K. Harada. 1967. Epidemiologic study of occupational lung cancer. *Arch. Environ. Health* 14, 859-864

Kennaway E.L. 1925. Experiments on cancer producing substances. *Br Med J.* 2: 1-4.

Kennaway, E.L. and Kennaway, N.M. 1937. *Acta Un.int.Cancer* 2:101.

Kennaway, E.L. and N.M. Kennaway. 1947. A further study of the incidence of cancer of the lung and larynx. *Br. J. Cancer* 1, 260-298

Kennaway E.L. and N.M. Kennaway. 1951. Studies of the incidence of cancer of the lung and larynx. *British Journal of Cancer* V: 153-158.

Kim, Y.H., D.T. Woodley, K.C. Wynn, W. Giomi, and E.A. Bauer. 1992. Recessive Dystrophic Epidermolysis Bullosa Phenotype Is Preserved in Xenografts Using SCID Mice: Development of an Experimental In Vivo Model. 0022-202X/92/\$05.00 Copyright 1992 by the Society for Investigative Dermatology, Inc.

Kroese, E.D., J.J.A. Muller, G.R. Mohn, P.M. Dortant, and P.W. Wester. 2001. Tumorigenic effects in Wistar rats orally administered benzo[a]pyrene for two years (gavage studies): Implications for human cancer risks associated with oral exposure to polycyclic aromatic hydrocarbons. (658603 010). Bilthoven, The Netherlands: National Institute for Public Health and the Environment (RIVM).

Kurtz, A., A. Aigner, R.H. Cabal-Manzano, R.E. Butler, D.R. Hood, R.B. Sessions, F. Czubyko, and A. Wellstein. 2004. Differential Regulation of a Fibroblast Growth Factor-Binding Protein during Skin Carcinogenesis and Wound Healing. *Neoplasia* Vol. 6: 595-602

Letzel, S. and H. Drexler. 1998. Occupationally related tumors in tar refinery workers. *J. Am. Acad. Dermatol.* 39, 712-720

- Levin, W., A.W. Wood, P.G. Wislocki, J. Kapitulnik, H. Yagi, D.M. Jerina, and A.H. Conney. 1977. Carcinogenicity of benzo-ring derivatives of benzo(a)pyrene on mouse skin. *Cancer Res* 37: 3356-3361.
- Levy, L. 1995. The 'Particle Overload' Phenomenon and Human Risk Assessment. *Indoor Environ* 4:254-262
- Lin, A.N. and K. Moses. 1985. Tar Revisited. *Int J Dermatol* 24, 216-218.
- Lindelof B. and B. Sigurgeirsson. 1993. PUVA and Cancer: A Case-Control Study. *Br J Dermatol* 129, 39-41.
- Liu, N., Z. Wang, D. Dong, K. Chen, and L. Quin. 1997. Cancer mortality among carbon workers in China: Retrospective cohort study. *J. Occup. Health* 39, 325–330
- Mackenna R.M.B. 1959. Uncomplicated Psoriasis. *Br Med J, Dec 1959* 244-1247
- Magee, B., K. Connor, D. Chin, and V. Houck. 2012. Critical Evaluation of USEPA's Toxicological Assessment of Benzo(a)pyrene. 2012 SETAC North America 33rd Annual Meeting. Long Beach, California.
- Martin, J.C., E. Imbernon, M. Goldberg, A. Chevalier, and S. Bonenfant. 2000. Occupational risk factors for lung cancer in the French electricity and gas industry. *Am. J. Epidemiol.* 151, 902–912
- Maughan W.Z., S.A. Muller, H.O. Perry, M.R. Pittelkow, and P.C. O'Brien. 1980. Incidence of Skin Cancers in Patients with Atopic Dermatitis Treated with Coal Tar. *Am Acad Dermatol* 3(6), 612-615
- Menter A. and D.L. Cram. 1983. The Goeckerman Regimen in Two Psoriasis Day Care Centers. *J Am Acad Dermatol* 9, 59-65.
- Merlo, D.F., S. Garattini, U. Gelatti, C. Simonati, L. Covolo, M. Ceppi, and F. Donato. 2004. A mortality cohort study among workers in a graphite electrode production plant in Italy. *Occup. Environ. Med.* 61, e9
- Milham, S. 1982. Occupational Mortality in Washington State 1950–1979, Prepared under Contract No. 210-80-0088, Cincinnati, OH, National Institute for Occupational Safety and Health
- Mohamed, E., W.H. Song, S.A. Oh, Y.J. Park, Y.A. You, S. Lee, J.Y. Choi, Y.J. Kim, I. Jo, and M.G. Pang. 2010. The transgenerational impact of benzo(a)pyrene on murine male fertility. *Hum Reprod* 25: 2427-2433.
- Moody, R.P., J. Joncas, M. Richardson, and I. Chu. 2007. Contaminated soils (I): In vitro dermal absorption of benzo[a]pyrene in human skin. *J Toxicol Environ Health A* 70: 1858-1865.
- Morabia A, S. Markowitz, K. Garibaldi, and E.L. Wynder. 1992. Lung cancer and occupation: results of a multicenter case-control study. *Br J Ind Med* 49: 721–727.

Mori, I. 2002. Cancer mortality among man-made graphite electrode manufacturing workers: Results of a 38 year follow up. *Occup. Environ. Med.* 59, 473–480

Morton, C.L., and P.J. Houghton. 2007. Establishment of human tumor xenografts in immunodeficient mice. *Natural Protocols* 2:247-250.

Moulin, J.J., P. Wild, J.M. Mur, M. Lafontaine, M. Lefer, M. Mercier-Gallay, P. Villemot, V. Whebi, and J.P. Coulon. 1989. Risk of lung, larynx, pharynx and buccal cavity cancers among carbon electrode manufacturing workers. *Scand. J. Work Environ. Health* 15:30–37.

Moulin, J.J., T. Clavel, B. Buclez, and G. Laffitte-Rigaud. 2000. A mortality study among workers in a French aluminium reduction plant. *Int. Arch. occup. environ. Health* 73:323–330

Muller S.A. and R.R. Kierland. 1964. Crude Coal Tar in Dermatologic Therapy. *Mayo Clin Proc* 39:275-280.

Muller S.A., H.O. Perry, M.R. Pittelkow, W.Z. Maughan, P.C. O'Brien. 1981. Coal Tar, ultraviolet Light, and Cancer. *J Am Acad Dermatol* 4(2):234-235.

Muller, S.A. and H.O. Perry. 1984. The Goeckerman Treatment in Psoriasis: Six Decades of Experience at Mayo Clinic. *Cutis* 34:265-269.

Mur, J.M., J.J. Moulin, C. Meyer-Bisch, N. Massin, J.P. Coulon, and J. Loulergue. 1987. Mortality of aluminium reduction plant workers in France. *Int. J. Epidemiol.* 16:257–264

Nesnow, S., L.L Triplett, and T.J. Slaga. 1983. Mouse skin tumor initiation-promotion and complete carcinogenesis bioassays: mechanisms and biological activities of emission samples. *Environ Health Perspect* 47: 255-268.

Nielsen. 1999. Nielsen Household Panel Distribution Report. Consumer stated usage of Denorex shampoo. As cited in ICF (2000).

Nomura, T., H. Nakajima, and T. Hongyo. 1997. Induction of Cancer, Actinic Keratosis, and Specific p53 Mutations by UVB Light in Human Skin Maintained in Severe Combined Immunodeficient Mice. *American Association for Cancer Research* 57:2081-2084.

Northcross, A.L., M. Trinh, J. Kim, I.A. Jones, M.J. Myers, D.D. Dempsey, N.L. Benowitz, and S.K. Hammond. 2012. Research Report: Particulate mass and polycyclic aromatic hydrocarbons exposure from secondhand smoke in the back seat of a vehicle. *Tobacco Control* 0:1-7.

NRC.2011. National Research Council. Review of the Environmental Protection Agency's draft IRIS assessment of formaldehyde. National Academies Press. 2011.

- NTP. 1993. National Toxicology Program. Toxicology and carcinogenesis studies of Talc in F344/N Rats and B6C3F1 mice. Technical Report Series No. 421. NIH Publication No. 93-315.
- NTP. 1996. National Toxicology Program. Comparative Initiation/Promotion Skin Paint Studies Of B6C3F₁ Mice, Swiss (CD-1®) Mice, And Sencar Mice. Technical Report Series No. 441.
- Oberdorster, G. 1995. The NTP Talc Inhalation Study: A Critical Appraisal Focused on Lung Particle Overload. *Reg Tox Pharm* 21:233-241.
- O'Donovan, W.J. 1920. Epitheliomatous ulceration among tar workers. *Br. J. Dermatol. Syphilis* 32: 215–228.
- OECD. 2004. *OECD Guideline for the Testing of Chemicals: Skin Absorption: in vitro Method*. OECD Test No. 428. The Organisation for Economic Co-operation and Development. April 13.
- OECD. 2010. *OECD Guidance Notes on Dermal Absorption*.
- Office of Science and Technology Policy. 1985. Chemical carcinogens: A review of the science and its associated principles. *Fed. Reg.*, 11: 10371-10442.
- Ova, G. and S. Onaran. 1998. Polycyclic aromatic hydrocarbons contamination in salmon-trout and eel smoked by two different methods. *Advances in food sciences* 20: 168-172. Franco. F., E. Chellini, and A. Costantini. 1993. Mortality in the coke oven plant of Carrara, Italy. *Med Lav*, 84: 443–447.
- Partanen, T. and P. Boffetta. 1994. Cancer risk in asphalt workers and roofers: Review and metaanalysis of epidemiologic studies. *Am. J. Ind. Med.* 26: 721–740
- Pauluhn, J., J. Thyssen, J. Althoff, G. Kimmerle and U. Mohr. 1985. Long-term inhalation study with benzo(a)pyrene and SO₂ in Syrian golden hamsters. *Exp Path* 28:31.
- Perry H.O., C.W. Soderstrom, and R.W. Schulze. 1968. The Goeckerman Treatment of Psoriasis. *Arch Dermatol* 98:178-182
- Phillips, D.H. 1983. Fifty years of benzo(a)pyrene. *Nature* 303:468-472.
- Pittelkow M.R., H.O. Perry, S.A. Muller, W.Z. Maughan and P.C. O'Brien. 1981. Skin Cancer in Patients With Psoriasis Treated With Coal Tar. *Arch Dermatol* 117:465-468.
- Poel, W.E. 1959. Effect of carcinogenic dosage and duration of exposure on skin-tumor induction in mice. *J Natl Cancer Inst* 22:19-43.

Poel, W.E. 1960. Skin as a test site for the bioassay of carcinogens and carcinogen precursors [IARC Monograph]. In IARC Monographs on The Evaluation of Carcinogenic Risk of Chemicals To Man (pp. 611-631). [Incorrect citation from USEPA (2013)]

Poel, W.E. 1963. Skin as a test site for the bioassay of carcinogens and carcinogen precursors. In National Cancer Institute Monograph No. 10; Conference on Biology of Cutaneous Cancer (pp. 611- 631).

Pott, P. 1775. *Chirurgical Observations Relative to the Cataract, the Polypus of the Nose, the Cancer of the Scrotum, the Different Kinds of Ruptures, and the Mortification of the Toes and Feet*. London: Hawes,Clarke & Collins, pp. 7–13.

Potter, D., E.D. Booth, H.C.A. Brandt, R.W. Loose, R.A.J. Priston, A.S Wright, and W.P. Watson. 1999. Studies on the dermal and systemic bioavailability of polycyclic aromatic compounds in high viscosity oil products. *Arch Toxicol* 73:129-140.

Pukkala, E. 1995. Cancer risk by social class and occupation: a survey of 109,000 cancer cases among Finns of working age. In J Wahrendorf (Ed.). (Karger: Basel, Switzerland).

Pukkala, E., J.I. Martinsen, and E. Lynge. 2009. Occupation and cancer - follow-up of 15 million people in five Nordic countries. *Acta Oncologica (Stockholm, Sweden)* 48: 646–790.

Randem, B.G., S. Langard, I. Dale. 2003. Cancer incidence among male Norwegian asphalt workers. *Am J Ind Med* 43: 88–95.

Reed, N.D. and D.D. Manning. 1973. Long-Term Maintenance of Normal Human Skin on Congenitally Athymic (Nude) Mice June 1, 1973. *Experimental Biology and Medicine Exp Biol Med* (Maywood) 1973 14: 350.

Reifenrath, W.G., E.M. Chellquist, E.A. Shipwash, W.W. Jederberg, and G.G. Krueger. 1984. Percutaneous penetration in the hairless dog, weanling pig and grafted athymic nude mouse: evaluation of models for predicting skin penetration in man. *British Journal of Dermatology (1984) III*, Supplement 27:123-135

Richmond, A. and Y. Su. 2008. Mouse xenograft models vs GEM models for human cancer therapeutics. *Dis Model Mech* 2008 Sep-Oct 1(2-3):78–82.

Rockett, H.E. and V.C. Arena. 1983. Mortality studies of aluminum reduction plant workers: Potroom and carbon department. *J. Occup. Med.* 25, 549–557

Roe, F.J., R. Peto, F. Kearns, and D. Bishop, D. 1970. The mechanism of carcinogenesis by the neutral fraction of cigarette smoke condensate. *Br J Cancer* 24: 788-806.

- Roelofzen, J., K. Aben, U. Oldenhof, P. Coenraads, H. Alkemade, P. van de Kerkhof, P. van der Valk, and L. Kiemeneij. 2010. No increased risk of cancer after coal tar treatment in patients with psoriasis or eczema. *Journal of Investigative Dermatology* 130: 953.
- Romundstad, P., A. Andersen, and T. Haldorsen. 2000a. Cancer incidence among workers in six Norwegian aluminum plants. *Scand. J. Work Environ. Health* 26:461–469
- Sakabe H., K. Tsuchiya, N. Takekura, S. Nomura, S. Koshi, K. Takemoto, H. Matsushita and Y. Matuso. 1975. Lung cancer among coke oven workers. *Ind. Health* 13, 57–68
- Schmähl, D., K.G. Schmidt, and M. Habs. 1977. Syncarcinogenic action of polycyclic hydrocarbons in automobile exhaust gas condensates. In *Air pollution and cancer in man*. Lyon, France: IARC.
- Schmidt, K.G., D. Schmähl, and J. Misfeld. 1973. Investigations on the carcinogenic burden by air pollution in man. VI. Experimental investigations to determine a dose-response relationship and to estimate a threshold dose of benzo(a)pyrene in the skin of two different mouse strains. *Zentralbl Bakteriol, Parasitenkd, Infektionskrankh Hyg, Abt 1: Orig, Reihe B* 158: 62-68.
- Schoenberg, J.B., A. Stemhagen, and T.J. Mason. 1987. Occupation and lung cancer risk among New Jersey white males. *J Natl Cancer Inst* 79: 13–21.
- Secretan, B., K. Straif, R. Baan, Y. Grosse, F. El Ghissassi, V. Bouvard, L. Benbrahim-Tallaa, N. Guha, C. Freeman, L. Galichet, and V. Coglianò. 2009. A review of human carcinogens--Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. *Lancet Oncol* 10: 1033-1034.
- Sim M.R., A. Del Monaco, and J.L. Hoving. 2009. Mortality and cancer incidence in workers in two Australian prebake aluminium smelters. *Occup Environ Med* 66: 464–470.
- Sivak, A., R. Niemeier, D. Lynch, K. Beltis, S. Simon, R. Salomon, R. Latta, B. Belinky, K. Menzies, A. Lunsford, C. Cooper, A. Ross, and R. Bruner. 1997. Skin carcinogenicity of condensed asphalt roofing fumes and their fractions following dermal application to mice. *Cancer Lett* 117: 113-123.
- Soballe PW, Montone KT, Satyamoorthy K, Nesbit M, Harlyn M. 1996. Carcinogenesis in human skin grafted to SCID mice. *Cancer Res* 56:757-764.
- Spinelli, J.J., P.A. Demers, N.D. Le, M.D. Friesen, M.F. Lorenzi, R. Fang, and R.P. Gallagher. 2006. Cancer risk in aluminum reduction plant workers (Canada). *Cancer Causes Control* 17: 939-948.
- Stern, R.S. and R. Lange. 1988. Non-melanoma skin cancer occurring in patients treated with PUVA five to ten years after first treatment. *J. Invest Dermatol* 91:120-124.
- Stern, R.S., S. Zierler, and J.A. Parrish. 1980. Skin carcinoma in patients with psoriasis treated with topical tar and artificial ultraviolet radiation. *Lancet* 1: 732-735.

Stern, R.S., E.J. Liebman, and L. Väkevä. 1998. Oral psoralen and ultraviolet-A light (PUVA) treatment of psoriasis and persistent risk of nonmelanoma skin cancer. PUVA Follow-up Study. *J Natl Cancer Inst* 90: 1278-1284.

Stern, F.B., A.M. Ruder, and G. Chen. 2000. Proportionate mortality among unionized roofers and waterprooferers. *Am. J. Ind. Med.* 37, 478–492

Straif, K., R. Baan, Y. Grosse, B. Secretan, F.E. Ghissassi, and V. Coglianò. 2006. Carcinogenicity of household solid fuel combustion and of high-temperature frying. *The Lancet Oncology* 7:977-978.

Stucker, I., D. Meguellati, and P. Boffetta. 2003. Cohort mortality study among French asphalt workers. *Am J Ind Med* 43: 58–68.

Swaen, G.M.H. and J.M.M. Slangen. 1997. Mortality in a group of tar distillery workers and roofers. *Int. Arch. Occup. Environ. Health* 70, 133–137

Swaen, G.M.H., J.J.M. Slangen, A. Volovics, R.B. Hayes, T. Scheffers, and F. Sturmans. 1991. Mortality of coke plant workers in the Netherlands. *Br. J. Ind. Med.* 48, 130–135

Teta, M.J., M.G. Ott, and A.R. Schnatter. 1987. Population based mortality surveillance in carbon products manufacturing plants. *Br. J. Ind. Med.* 44, 344–350

Thyssen, J., J. Althoff, G. Kimmerle, and U. Mohr. 1981. Inhalation studies with benzo[a]pyrene in Syrian golden hamsters. *J Natl Cancer Inst* 66: 575-577.

Torinuki W. and H. Tagami. 1988. Incidence of Skin Cancer in Japanese Psoriatic Patients Treated with Either Methoxsalen Phototherapy, Goeckerman Regimen, or Both Therapies. *J Am Acad Dermatol* 18, 1278-1281.

Törnqvist, S., S. Norell, A. Ahlbom, and B. Knave. 1986. Cancer in the electric power industry. *Br J Ind Med* 43: 212-213.

Urano, K., Y. Katakai, Y. Tokuda, Y. Ueyama, T. Nomura, and S. Yamamoto. 1995. Failure of genotoxic carcinogens to produce tumors in human skin xenografts transplanted to SCID mice. *Carcinogenesis* 16: 2223-2226.

USEPA. 1988a. Recommendations for and documentation of biological values for use in risk assessment. (USEPA/600/6-87/008). Cincinnati, OH.

USEPA. 1988b. Summary of the Second USEPA Workshop on Carcinogenesis Bioassay via the Dermal Route. USEPA/600/6-89-003. Research Triangle Park, North Carolina. May.

USEPA. 1989. *Risk Assessment Guidance for Superfund, Volume I, Human Health Evaluation Manual, Part A*. Office of Emergency and Remedial Response, Washington, D.C. USEPA/540/1-89-002. December.

USEPA . 1990. Development of a dose-response model for inhaled B(a)P. (USEPA-600/R-92-135).

USEPA. 2000a. *Benchmark Dose Technical Guidance Document*. USEPA/630/R-00/001. Draft. U.S. Environmental Protection Agency, Risk Assessment Forum, Washington, DC. October.

USEPA. 2000b. Science policy council handbook: Risk characterization [USEPA Report]. (USEPA 100-B-00-002). Washington, D.C.

USEPA. 2003. *A Summary of General Assessment Factors for Evaluating the Quality of Scientific and Technical Information*. U.S. Environmental Protection Agency.

USEPA. 2005. Guidelines for Carcinogen Risk Assessment . (USEPA/630/P-03/001F). Washington, D.C.

USEPA. 2007. *Guidance for Superfund Volume I: Human Health Evaluation Manual (Part E, Supplemental Guidance for Dermal Risk Assessment)*. Final Revision. October.

USEPA. 2010. *Multistage Weibull time-to-tumor model in USEPA's benchmark dose software (BMDS): Methodology description*. Revised draft: Washington, DC.

USEPA. 2012. *Regional Screening Levels. Users Guide*. November.

USEPA. 2013. *Toxicological Review of Benzo[a]pyrene (CASRN 50 32 8) In Support of Summary Information on the Integrated Risk Information System (IRIS)*. USEPA/635/R13/138a. Public Comment Draft. U.S. Environmental Protection Agency, National Center for Environmental Assessment, Office of Research and Development, Washington, DC. August.

USGS. 2003. Concentrations of Polynuclear Aromatic Hydrocarbons and Inorganic Constituents in Ambient Surface Soils, Chicago, Illinois: 2001-02. Water-Resources Investigations Report 03-4105. In cooperation with the Chicago Department of Environment.

Van Schooten, F. and R. Godschalk. 1996. Coal Tar Therapy: Is it Carcinogenic? *Drug Safety* 15(6):374-377

Wei, S.J., R.L. Chang, K.A. Merkle, M. Gwynne, X.X. Cui, B. Murthy, M.T. Huang, J.G. Xie, Y.P. Lu, Y.R. Lou, D.M. Jerina, and A.H. Conney. 1999. Dose-dependent mutation profile in the c-Ha-ras proto-oncogene of skin tumors in mice initiated with benzo[a]pyrene. *Carcinogenesis* 20: 1689-1696.

Wester, R.C. and H.I. Maibach. 1989. In Vivo Animal Models for percutaneous Absorption. In Bronasugh, R.L. and Maibach, H.I. Percutaneous Absorption (Marcel Dekker: New York). Pages 221-238.

- Wong, O. and F. Harris. 2005. Retrospective cohort mortality study and nested case–control study of workers exposed to creosote at 11 wood-treating plants in the United States. *J. Occup. Environ. Med.* 47, 683–697
- Wu, W. 1988. Occupational cancer epidemiology in the People's Republic of China. *J. Occup Med.* 30, 968–974
- Wu-Williams, A.H., Z.Y. Xu, W.J. Blot, X.D. Dai, R. Louie, H.P. Xiao, B.J. Stone, X.W. Sun, S.F. Yu, and Y.P. Feng. 1993. Occupation and lung cancer risk among women in northern China. *Am. J. ind. Med.* 24, 67–79
- Wu, J., A. Ramesh, T. Nayyar, and D.B. Hood. 2003. Assessment of metabolites and AhR and CYP1A1 mRNA expression subsequent to prenatal exposure to inhaled benzo(a)pyrene. *Int J Dev Neurosci* 21: 333-346.
- Xu, Z., L.M. Brown, G.W. Pan, T.F. Liu, G.S. Gao, B.J. Stone, R.M. Cao, D.X. Guan, J.H. Sheng, Z.S. Yan, M. Dosemeci, J.F. Fraumeni, and W.J. Blot. 1996. Cancer risks among iron and steel workers in Anshan, China, Part II: Case-control studies of lung and stomach cancer. *Am J Ind Med* 30:7-15.
- Xu, C., J. Chen, Z. Qiu, Q. Zhao, J. Luo, L. Yang, H. Zeng, Y. Huang, L. Zhang, J. Cao, and W. Shu. 2010. Ovotoxicity and PPAR-mediated aromatase downregulation in female Sprague-Dawley rats following combined oral exposure to benzo[a]pyrene and di-(2-ethylhexyl) phthalate. *Toxicol Lett* 199: 323-332.
- Zahm, S.H., R.C. Brownson, J.C. Chang, and J.R. Davis. 1989. Study of lung cancer histologic types, occupation, and smoking in Missouri. *Am J Ind Med* 15: 565–578.
- Zaidi, S.N.H., I. Laidlaw, A. Howell, C.S. Potter, D.P. Cooper, and P.J. O'Connor. 1992. Normal human breast xenografts activate N-nitrosodimethylamine: identification of potential target cells for an environmental nitrosamine. *Br J. Cancer* 66, 79-83
- Zheng, S.J., H.J. Tian, J. Cao, and Y.Q. Gao. 2010. Exposure to di(n-butyl)phthalate and benzo(a)pyrene alters IL-1 β secretion and subset expression of testicular macrophages, resulting in decreased testosterone production in rats. *Toxicol Appl Pharmacol* 248: 28-37.