Appeals, 5th Circuit, in New Orleans. However, the publicity surrounding CPSC's efforts to ban UFFI, and its assessment of the potential harm due to formaldehyde off gassing from this product, led to a severe reduction of the residential use of UFFI at the time. Subsequent efforts such as the creation of an informational brochure for consumers (1997) have attempted to further reduce exposure to exogenous formaldehyde.

The FA IRIS document is comprised of 4 volumes; 1) Introduction, Background, and Toxicokinetics, 2) Hazard Characterization, 3) Quantitative Assessment: Inhalation Exposure and Major Conclusions and 4) Appendices A-H. Approximately 900+ pages of documentation describe the stages for estimation of multiple inhalation reference concentrations (RfCs) and unit cancer risks based on different toxicological pathologies.

CPSC staff is pleased to have had the opportunity to review and provide comments on the FA IRIS document. Review of the FA IRIS document will take place in order of discussion in the document. Scientific commentary will be provided for many areas. Commentary on policy may also be noted, since science and policy is sometimes blurred (i.e. selection of uncertainty factors), even though EPA has remarked that these comments will not be considered in stakeholder response reports. Finally, a lack of commentary on issues presented in the FA IRIS document does not necessarily indicate agreement with what has been written. The substantial size of the document and the need for accelerated review (< 30 days) has precluded in-depth review for a significant portion of the document.

Overall Commentary

This document has done a remarkable job of summarizing and evaluating an immense amount of data on the potential toxicity and risks following exposure to formaldehyde. In general, the text is well written and understandable throughout. There are, however, a few issues that detract from the overall conclusions presented. These are presented below.

1) A primary issue with the FA IRIS document is that the discussion of study limitations (i.e. small number of test subjects, confounding-related issues, methodological/statistical issues) seems inconsistent in many cases. Some studies receive substantial criticism and others receive little or none without qualifying statements as to why. Inconsistent application of study criticisms gives the appearance of partiality.

2) The FA IRIS document is also very difficult to read in its current form. Document readability can be improved substantially by considering the following:

A) Merge the 4 volumes in to 1 volume. This will eliminate the repetitious nature of seeing the same Contents, List of Tables, List of Figures, and List of Abbreviations and Acronyms for each volume (removing approximately 90 pages of repetition).

B) For the main document, use a consecutive numbering scheme not based on Chapters (i.e. use 1, 2, 3 instead of 1-1, 1-2, 1-3)

C) Convert measures of atmospheric concentration to a standard unit such as ppb, ppm, or $\mu g/m^3$.

D) Italicize et al. to *et al.*

E) Data for some studies seems to be segregated into a variety of places in the document (i.e. Malek et al. 2003a; 4-256, 257; 5-23). The FA IRIS document would be easier to read and understand if all of the pertinent data or methods were contained in one place and then referred back to by other sections when needed.

3) At least one internal referral (i.e. to section 4.2.1.4, Page 5-18) is inaccurate (should be to 4.2.1.6). The authors should check other internal referrals to ensure that these are correct.

4) In Point of Departure Tables (i.e. Table 5.1, 5-3, etc), it is important to note additional data such as the number of test and control subjects (or litters, i.e. "n"), the statistical significance of the response (i.e. $P<0.05$), and/or the biological significance of the response (i.e. 10% decrease).

5) It is unclear to the reviewer what the selection criteria were for studies included in Point of Departure Tables. For example, in the case of neurological and behavioral POD tables, some species (mice) were excluded on the basis of reflex bradypnea, while studies with similar exposures to rats were included. The Table also included results from a human study with numerous confounders and other deficiencies, which were seemingly substantial enough to warrant exclusion from the Table. Clarifying the criteria for inclusion in the POD tables (i.e. using only "acceptable" studies) would go a long way to resolving confusion associated with their subsequent interpretation. Note that disqualification from the POD Tables could be based on study deficiencies. Subsequent exclusion from RfC derivation (i.e. the Pitten et al (2000) study, Page 5-22) should be based on other criteria (i.e. inability to extrapolate exposure durations).

6) On some of the study summaries (i.e. page 4-255, line 14, etc); the exposure duration is left out of the description. This is an important piece of methodological information that should be provided in order to judge durational effects.

7) In the portions of the FA IRIS document reviewed by CPSC staff, there were no incidences in which a review of publications resulted in the conclusion that there were not enough consistent information/data to conclude with any certainty that the effect occurred. EPA should consider the addition of this possibility to discussions related to adverse effects.

8) Overall, unrelated pathologies should not serve as mutual support for the conclusion that an effect on a particular organ system is occurring. If discrete pathologies (i.e. change in

fecundity) are not consistent among studies, effects should be discounted or caveated in discussion material.

9) Overall, a variety of study endpoints (i.e. pathologies) have been subjectively deemed significant, either in disregard for the statistics that demonstrate otherwise, or using the argument of biological plausibility. A lack of statistical association generally means either 1) there is no association, or 2) there are study confounders (i.e. group size) that preclude the generation of an association. Making the jump to biological plausibility from non-significant data should be discouraged except in special circumstances in which these specific pathologies are also demonstrated in other studies at similar doses.

Volume I discusses the properties, uses, environmental concentrations, human exposures, protein binding, endogenous and exogenous sources, metabolites, absorption, metabolism, distribution, excretion, and the toxicokinetics of formaldehyde and DNA-protein crosslinks.

No comments on this section.

Volume II is a hazard characterization and reviews data on human and animal non-cancer effects (i.e. sensory irritation, lung function, asthma, neurological, developmental and reproductive) and cancer effects (respiratory, gastrointestinal, lymphohematopoetic) effects. It also covers genotoxicity (i.e. formaldehyde-DNA reactions, *in vitro* clastogenicity, *in vitro* mutagenicity, *in vivo* mammalian genotoxicity), potentially susceptible populations, co-exposures, and database uncertainties.

General Commentary

1) Review of many of the inhalation studies did not include a description of the physical characteristics of the experimental dose (i.e. mass median aerodynamic diameter, MMAD; particle droplet size distribution). Without this information, it is impossible to determine where the dose is being deposited (i.e. nasopharyngeal area, bronchi, bronchioles, alveoli), and hence determine the relevance or lack of pathologies. Inhalation studies should include a physical characterization of the exposure aerosol or discuss potential complications with not having this information (in terms of pathologies seen or not seen). Grading the studies (as to their confidence; low, medium, high) should also consider the lack of this information.

Specific Commentary

Page 4-44, line 1-17 – Were the dose levels (atmospheric concentrations) measured or inferred for these various chemicals? If this study is to be used as primary support for suggesting neurological impairment ("…and the epidemiology study by Weisskopf et al. (2009) provide strong support for an association between increasing neurotoxicity and increasing duration of exposure"), more study details should be provided.

Page 4-45, line 18-19 – Suggest deleting "There is….needed." because the "serious concern" generated by epidemiological and controlled exposure human studies assessing neurobehavioral endpoints may be overstated. As stated on Page 4-42, line 31, "some suggestion of neurological impairment" is probably a better descriptor, since the studies are really a mixed bag of possibly positive, equivocal, and negative results. Perhaps this sentence could be worded something like "Additional research is needed to address these limitations".

Page 4-257, Figure 4-21 – There are no significance or error bars assigned to any of the treatment groups in this Figure. Are any of these results significant?

Volume III covers the quantitative assessment of RfCs from inhalation exposures. Included in this section are candidate RfCs for: 1) sensory irritation of the eyes, nose, and throat; 2) upper respiratory tract pathology; 3) pulmonary function effects; 4) asthma and allergic sensitization; 5) immune function; 6) neurological and behavioral toxicity, and; 7) developmental and reproductive toxicity. A quantitative cancer assessment (based on the NCI cohort study) is also provided for 1) nasopharyngeal cancer, 2) lymphohematopoetic cancer (Hodgkin lymphoma and Leukemia). The cancer assessment is supported by dose-response modeling of animal squamous cell carcinoma in the respiratory tract. Inhalation unit risk estimates are estimated for both human and animal data.

General Commentary

1) The broad application of a database uncertainty factor to account for reproductive and/or developmental effects is unwarranted when assessing RfCs (or RfDs) for various target organs. A database uncertainty factor should be applied to a particular target organ endpoint only if data are lacking for, or are not able to be extrapolated to, that endpoint. For example, if a hypothetical RfC was based on coronary effects, an extra database uncertainty factor for reproductive data deficiencies might not be applied. If, however, developmental data were lacking and developmental processes in the heart could potentially be affected, then a database uncertainty factor for developmental affects might be applied.

Specific Commentary

Page 5-2, line 10 – Two commas after category

Page 5-3, line 30 – This sentence makes the process of evaluating whether the study might have been influenced by reflex bradypnea sound very subjective. How is it possible to differentiate whether there is 1) no effect on an organ system because it is not affected by formaldehyde or 2) no effect on an organ system because reflex bradypnea decreased the overall systemic dose?

Page 5-6, line 19 – The sentence starting with "The mean score" is awkward. Also, should the statistical measure be the geometric mean instead of the mean?

Page 5-6, line 24 – The lack of correlation between duration of exposure and histologic changes decreases the confidence one has in using this study for RfC derivation.

Page 5-7, line 21 – How many students and is the difference in PEF biologically or statistically significant? What is the variation of these measures?

Page 5-8, line 5 – "chrinic" after "and"

Page 5-9, line 28 – Is the difference statistically significant? Can it be distinguished from normal variation?

Page 5-10, line 14-21 – What relevance do dermal studies have in a discussion of inhalation-induced pathologies? Either qualify why it is OK to evaluate dermal studies (i.e. because the immune pathways for hypersensitization are the same in dermal and respiratory systems) or support this section with examples of hypersensitization derived from inhalation studies.

Page 5-11, line 17 – How does the reported question compare to a medical diagnosis of asthma? What other diseases have the same reportable symptoms ("attack", "wheezing", "short of breath"), and in what comparative frequency?

Page 5-11, line 21 – Indoor air concentrations of formaldehyde (2.3 $\mu g/m^3$) are less than outdoor air concentrations (5.8 $\mu g/m^3$)? Is this correct? Indoor air concentrations in this study are substantially below all levels cited in other review documents (i.e. NTP ROC, 2010, page 60-63).

Page 5-11, line 22-24 – Cumulative asthma and daytime attacks of breathlessness were associated with outdoor exposure levels (4.7 ppb), but not with indoor exposure levels (1.9 ppb)? Were these values distinguishable statistically or biologically?

Page 5-13, line 15-17 – Please explain why a "no effect" determination is "consistent with two-fold increase in risk"

Page 5-13, line 30-33 – This sentence is unclear. Is the author using peak measurements as exposure estimates for each group?

Page 5-14, line 3-4 – OR's of 1.34 and 1.42 are very weak positive associations.

Page 5-14, line 9-12 – As with line 30-33 of the previous page, the meaning of this sentence is unclear.

Page 5-14, line 25-26 – An OR of 1.39 is a very weak positive association.

Page 5-14, line 30-33 – What is the citation? How were the other confounders in cigarette smoke controlled for? Why are the concentration cutoffs 40 and 60 ppb? What is the variation around the percentages of asthma? A dose-response is not evident for this study (<40 ppb =

15.1%, 40-60 ppb = 0%, >60 ppb = 45.5% asthma). Trends with non-smokers (range 4.3-14.0 ppb) seem to refute positive asthma findings at lower concentrations (page 5-11, line 22-24).

Page 5-14, line 34 – Delete blank line

Page 5-15, line 2 – What was the geometric mean for formaldehyde concentrations? and were the average concentrations significantly different?

Page 5-16, line 9, line 21 – Use of the word "potentiation" implies a synergistic response rather than just an additive one. Is this what the author meant?

Page 5-16, line 19-20 – Data exists to support a neurogenically-mediated hypersensitization response rather than an immunological response. Should these two different types of responses be separated into different categories since their mode of actions are different?

Page 5-17, line 14 – Is there any supporting animal information for decrements in immune function? The small size of the supporting human study (line 7) precludes the use of this as a sole support for RfC derivation.

Page 5-17, line 9-10 – Addition of the erythrocyte count and hematocrit discussion does not add support to decrements in immune function, since these are not functionally related to the immune system.

Page 5-17, line 23 – What does "inform the formaldehyde RfC" mean?

Page 5-17, line 35 – Mice were not considered for RfC development because of potential reflex bradypnea? The doses used in mouse studies (Table 4-57) are substantially similar or lower than doses used in similar rat studies. Using this argument, why aren't the high dose rat studies also eliminated. Is this criteria considered in the evaluation of other pathological endpoints (i.e. are all studies with potential reflex bradypnea discounted in POD tables?).

Page 5-18, line 1 – reference to Section 4.2.1.4 refers to "Immune Function" section, and should refer to "Neurological and Neurobehavioral Function" section (4.2.1.6).

Page 5-19, Table 5-1 – As commented upon on page 4-44 and 4-45, Bach et al. considered their results preliminary until confirmed. Other issues such as a small number of test subjects, no observable symptoms of respiratory irritation following doses up to 1.1 ppm, non-dose related effects, and no adjustment for numerous confounders, suggests that this study should not be selected for consideration in the POD table (under "Human neurobehavioral outcomes"). Other issues with this study are noted on Page 5-22, line 1-13.

Page 5-19, Table 5-1 – There seems to be a discrepancy in the reporting of exposure duration and other factors for Bach et al. (1990; as described in 4.1.1.6.2, page 4-44) in the Table (Table 5.1) and text (page 5-21, line 5).

Page 5-21, Figure 5-1 – The Figure contains the unit $(mg/m^3)^1/_3$. Is this correct? Also, calculated concentrations appear to be 260, 471, 634, and 813 ppb, rather than the 32, 170, 390, or 890 ppb displayed in the title.

Page 5-23, Table 5-2 – Swimming times are significantly different than controls at a LOAEL of 0.5 ppm for male and female rats (Day 6 and 10). When the data is normalized to control values the same dose-dependent relation is seen. Overall, the effect is mitigated substantially as time following dose increases. Significant increases in the swimming time error rate are also observed at a LOAEL of 0.1 ppm. Both the raw and normalized data do not display a dose-response relationship. As with swimming time, the error rate is mitigated over time when comparing Day 6 to Day 10 exposures. Mitigation of the higher swim times and error rates in a short amount of time (with continued dosing) suggest that adaptation, accommodation, or compensation is occurring. Since the observed effects are substantially reversible in a short amount of time, and the effects are not dose related, choice of this data set to use in RfC calculation is not warranted, in our view.

Page 5-23, line 9-16 – Methodologies for the swim time/water maze testing in Malek (2003a,c) are not provided in sufficient detail to facilitate a rigorous review of these studies (or in 4-255-4-258).

Page 5-24, Figure 5-2 – Normalization to Day 1 results may remove the bias inherent in displaying raw data.

Page 5-24, line 11 – Concentration related effects were seen for swimming time and normalized swimming time, not mean error rates.

Page 5-25, line 3-24 – Behavioral studies are not supported by neural pathology endpoints. Issues associated with selection of the Malek studies as an RfC have been discussed above.

Page 5-26, line 28 – extra space prior to “Of”.

Page 5-27, line 1 – Was cumulative exposure reported for this study, and if so, was it significant?

Page 5-27, line 6 – Increased ORs do not appear to be dose related among the low, medium, and high exposure groups, weakening effect arguments for this study. Further, the range of ORs for the high exposure group (1.2-8.3) is twice as high as that for the low (1.2-4.8) and medium (0.8-4) groups. Was the increased OR in the high dose group caused by a high outlier?

Page 5-27, line 8 – The use of “lost statistical significance” implies that the gloved workers were significantly different than controls. If this is accurate, please state this. From Page 5-30, line 5, this may not be accurate?

Page 5-27, Figure 5-3 – The 95% CIs in this figure appear to be asymmetrically distributed around each Fecundity Density Ratio center point.

Page 5-28, line 18-21 – The lack of additional doses and repeatability in other studies should preclude the use of the Senichenkova studies in POD tables. As with that defined above, criteria used to define POD tables for developmental and reproductive effects have not been defined. Studies with significant deficiencies (large uncertainties) should not be presented in POD tables.

Page 5-30, line 5 – "Although this FDR is not statistically significant, it can reasonably be assumed to be part of a trend of decreased FDR with increasing inhalation exposure". The jump from non-significance to biological plausibility should only be considered in circumstances in which these discrete pathologies are repeated in other studies at similar dose levels.

Page 5-44, line 30 - Choice of the NOAEL for the asthma POD (40 $\mu g/m^3$) was based on a "biologically significant" trend of increasing asthma risk. This exposure dose had a non-significant Odds Ratio (OR) of less than one. The next highest exposure dose (50-59 $\mu g/m^3$) was also had a non-significant OR of approximately 1.2 (estimated LOAEL). The high dose (60+ $\mu g/m^3$) was significantly different from controls and had an OR of 1.39 (weak association). Choice of a non-significant exposure dose with a weak association to asthma (low OR) to represent a LOAEL is unwarranted, in our view (see comment #9 above). Further, the use of the exposure group 60+ $\mu g/m^3$ as a LOAEL is problematic because the grouping does not identify the range or average of exposures included. Because of this, the NOAEL can't be used as a point of departure with confidence (i.e. is the LOAEL 60, 600, 6000 $\mu g/m^3$ and if it's 6000, is a NOAEL of 54 $\mu g/m^3$ a good POD to use?). Another option EPA should consider is to use the dose-response data and perform a benchmark dose-type of approach to estimate a nominal POD.

Page 5-74, line 29 and on - Most of the evidence used to support an association between formaldehyde and non-respiratory tract cancers (specifically all LHP, myeloid leukemia, all leukemia, multiple myeloma and Hodgkin lymphoma) comes from the NCI cohort. EPA's statements that "Epidemiologic studies involving formaldehyde-exposed workers provide sufficient evidence of a causal association between formaldehyde exposure and all LHP malignancies" (4.5.2.2), "The associations between myeloid leukemia and formaldehyde exposure are strong and 14 consistent" (4.5.2.5), and "There is evidence for an exposure response relationship for both Hodgkin lymphoma and multiple myeloma in the NCI industrial cohort among exposed workers" (4.5.2.6) would be much more difficult to make without the consideration of the many analyses of that cohort. EPA chooses this cohort as a basis for risk estimates for these endpoints for this reason and many others (including the strength of the exposure data) (5.2.1), and it is agreed that if a study had to be chosen for this purpose, the NCI cohort would be most relevant.

However, with one exception (Hodgkin lymphoma), all of the conclusions from the latest analysis of the NCI cohort (Beane-Freeman et al 2009) are based on peak exposure only. The evidence from average exposure, and the most important determinant, cumulative exposure, show no significant associations, either at any given exposure range or overall exposure-related trend, between exposure and any of the endpoints except for Hodgkin lymphoma. And for Hodgkin lymphoma, this only applies for average exposure, and not, again, for the most important determinant, cumulative exposure.

EPA chose to provide estimates of risk based on all leukemia and Hodgkin lymphoma. For all leukemia, there is nothing significant at all based on cumulative or average exposure. For peak exposure, there is a significant exposure-related trend when considering both unexposed and exposed person-years.

The problem, in our view, is the usefulness of "peak exposure". The authors themselves state "Peak exposures could be related to either routine or non-routine high-exposure tasks or could result from non-routine unusual events, such as spills. Peak frequency was estimated on the basis of knowledge of the job and the likelihood that a high-exposure task or event would occur. Because there was no direct information on frequency of peaks, a higher degree of uncertainty is associated with the frequency than the level of peak exposure"; peak exposures were also defined as short-term exposures, generally less than 15 minutes. Given this large uncertainty, and the lack of significance with any other metric, it seems premature at this time to use this as a basis of cause and effect, and then to take the non-significant exposure trend for cumulative exposure and base an estimate of risk on it.

For Hodgkin lymphoma, there are significantly elevated relative risks at the two highest peak exposure ranges, and the overall trend is significant compared to the referent (low) exposure level. The same is true for average exposure, but the RR at the highest level is lower than that at the second highest level. Again, for cumulative exposure, the RR at the highest level is lower than that at the second highest level, but nothing is significant.

The problems with peak exposure are the same as described for all leukemia. The average exposure suffers from two uncertainties: first, it does not consider number of years of exposure (a critically important parameter – which is why cumulative exposure is used) and second, the relative risk at the higher dose is lower than that at the second highest exposure, which is worrisome.

Again, it seems premature at this time to use this as a basis of cause and effect, and then to take the non-significant exposure trend for cumulative exposure and base an estimate of risk on it.

Page 6-2, line 4-5 – formaldehyde what? Decreased and where was this measured?

Page 6-6, line 23 – add "allergic" before rhinitis?

Volume IV contains the Appendices. These Appendices cover external peer review and public comments, modeling simulations, a lifetable analysis, sensitivity analyses, and an expert panel consultation (on the use of animal data for estimating quantitative cancer risk).

No comments

Thank you for the opportunity to review and comment on this draft document. If you have any questions, please feel free to contact me (kcarlson@cpsc.gov).

Sincerely,

Kent Carlson, Ph.D.

Toxicologist, Health Sciences Directorate