Preliminary Comments from Members of the CAAC Augmented for the Ethylene Oxide Review on EPA's Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide (Revised External Review Draft - August 2014).

Received as of 11-14-14

Dr. Henry A. Anderson 2
Dr. James V. Bruckner 3
Dr. W. Michael Foster 6
Dr. Gary Ginsberg 10
Dr. Steven Heeringa 13
Dr. Lawrence H. Lash 16
Dr. Maria Morandi 20
Dr. Victoria Persky 24
Dr. Kenneth Ramos 26
Dr. Lianne A. Sheppard 27
Dr. Daniel Zelterman 35

Dr. Henry A. Anderson

CQ6 (Appendix J – New studies)

The description of the search protocol is straight forward and easy to follow. It would be useful to mention if each potential reference was actually read and reviewed by someone or was only the abstract used to determine relevance. The categorization of the references into categories such as methods, exposure studies, etc. was useful. Table J-1 provides appropriate disposition statements and J2 is also a concise summary. A quick review of some of these papers determined that the groupings were appropriate.

The reviews of the four major studies provided sufficient information to defend EPA's decisions. Format wise, it would have been useful if EPA had separated the description of the studies and their key findings from the EPA analysis of the paper from an IRIS utility perspective. It is somewhat difficult to separate what was reported in the manuscript from what was EPA analysis and conclusions. For instance on page J-5 line 25 there is a discussion of low prevalence of exposure in the study and impact on study power. Is this from the authors or an EPA conclusion? The comment is appropriate but it would be useful to consider having EPA's interpretation/comments labeled as a section in the review. Another example would be page J-7 line 15. Is the comment about internal and external analyses from the authors or are these EPA determinations. I agree with the statements but it would help the reader to separate what the authors said from how EPA interpreted the results from the perspective of IRIS needs. Page J-8 line 6 on appear to be EPA conclusions. EPA may want to consider how to separate the manuscript description from the EPA assessment component.

It is a challenge to determine where the Valdez-Flores and Sielken (2013) paper (Section J.3.1) belongs. Because it is a critique of the EPA methods and the earlier version of the EtO carcinogenicity risk assessment it is an important document for EPA to review and comment upon, but is this a new study? In any case, this specific review may deserve a different organizational format. While it is usually clear when EPA is assessing components of the manuscript, the statements from the article tend to flow into the EPA comments/response.

EPA is to be commended for updating the literature search and including this section in the updated assessment.

Dr. James V. Bruckner

Charge Question 5b

Several SAB (2007) panel members strongly advocated considering using both linear and non-linear approaches for low-dose extrapolation, but EPA decided that inclusion of a non-linear approach was not warranted. In its rationale for this decision, EPA relied heavily upon its established policy on this topic. There was relatively little reliance on, or reference to published scientific findings on the subject. A statement that "additivity to background carcinogenic processes at low doses is expected to result in incremental risk that approaches linearity" was quoted and attributed to Crump et al. (1976). It was also noted that recent data of Marsden et al. (2009) support a linear exposure-response relationship for EtO exposure and DNA adducts. Marsden et al. presented dose-response data for exogenous N7-HEG adduct formation in 3 tissues of rats dosed with EtO. Modeling revealed that the response data best fit a linear model. The researchers emphasized, however, that the study was not designed to test for linearity, and that their data may not have been sufficient to allow establishment of a linear dose-response relationship. They recommended inclusion and analysis of 4 more low EtO doses, to confirm linearity or non-linearity. Their current plots of exogenous adducts versus EtO dosage-level exhibit what appear to this reviewer to be the suggestion of hockey-stick shapes.

The EPA fails to include an important conclusion of Marsden et al. (2009), namely that the exogenous (EtO-induced) DNA adducts in their rat tissues "constitute a relatively small proportion (<2%) of N7-HEG adducts at EtO doses < 0.005 mg/kg, and can be considered negligible". The researchers go on to say that "given the inherent variation in endogenous adduct levels in rats, it is important to recognize that the increase in adducts caused by direct binding of [^{14}C] EtO to DNA is not significant at any but one high-dose in one tissue. EPA fails to mention these important conclusions.

I would urge the EPA to make an effort to be more balanced in its presentation of the pros and cons of the conclusion that mutagenicity is the sole mechanism of EtO carcinogenicity. The EPA does emphasize, in its expanded discussion of the mutagenic mode of EtO action, that cancers are a consequence of an accumulation of genetic and epigenetic changes involving multiple genes and chromosomal aberrations. It is also noted that N7-HEG, the predominant EtO-DNA adduct, is unlikely to be promutagenic but could be subject to depurination, resulting in an apurinic site that could be vulnerable to miscoding during cell replication. Rusyn et al. (2005) presented evidence that does not support this as a potential mechanism for subsequent DNA strand breaks, as the DNA damage in brain, liver and spleen was apparently repaired without accumulation of such sites. EPA argues that their conclusions may have been different had they examined still another cell type. Tomkins et al. (2009) reported that plasmid mutation frequencies were significantly elevated only at very high EtO concentrations. EPA countered that

increased mutation frequency could not be ruled out in light of limitations in the study and variability of the data. Kirman et al. (2004) proposed there was a quadratic relationship between EtO exposure and leukemia risk, based on the premise that chromosomal aberrations (rather than gene mutations) are the characteristic initiating event for leukemia. EPA, in response, stated that chromosomal aberrations could be a downstream event, and that evidence was inadequate to support the researchers' premise. EPA does concede that evidence of EtO-induced mutation in humans is limited. The objectivity of EPA's EtO document would be improved by giving a more balanced interpretation of scientific findings pertinent to mechanism(s) and potential mechanism(s) of action, rather than merely supporting their conclusion that EtO is unquestionably a direct acting mutagen.

EPA should consider and discuss key issues/points raised by recognized authorities on the mode of carcinogenic action and linearity/non-linearity of pertinent effects of EtO. Swenberg et al. (2011), for example, published a comprehensive review of endogenous and exogenous DNA adduct formation, DNA repair, mutagenesis and carcinogenesis of formaldehyde, vinyl chloride and EtO. They raise some of the same questions that I have in my preceding comments. They also point out that the numbers of N7-HEG adducts measured in rat tissues by Marsden et al. (2009), in responses to the lowest EtO doses, were 1,000- and 10,000-fold lower than endogenous N7-HEG levels found in normal/control rodent and human tissues, respectively. This suggests that exogenous (EtO-induced) N7-HEG formation does not pose additional risk above that posed by ubiquitous background damage. Such findings indicate that EtO is a weak mutagen, which is consistent with observations in epidemiology studies of low cancer incidence, even in the most highly-exposed workers. Apparently, authoritative reviews and such conclusions by Swenberg et al. (2011) and others were "not considered further" by EPA.

Charge Question 6

The authors of Section J have done a good job presenting and discussing the results of the epidemiology studies of ethylene oxide (EtO)-exposed workers by Kiran et al. (2010) and Mikoczy et al. (2011). The descriptions of these investigations and the results are clear and appear to be accurate. The documents authors' interpretation and summations of the findings are transparent and reasonable, as are the accounts of each study's strengths and limitations.

Inclusion of the studies of Kiran et al. (2010) and Mikoczy et al. (2011) would not alter the hazard assessment of EtO. They do add to the weight-of-evidence of a causal relationship between EtO exposure and lymphoreticular (LR) and breast cancer in humans. I do not believe, however, that they support the conclusion that EtO is a strong (but less than conclusive) human carcinogen. Mikoczy et al. (2011) for example, only reported nonsignificant increases in SMRs and SIRs for LR cancer. Breast cancer SIRs were nonsignificantly decreased, although incidence rate ratios were significantly increased for the most highly-exposed individuals.

The authors of section J have written very comprehensive accounts of the findings and arguments of Valdez-Flores and Sielken (2013) and of Parsons et al. (2013). EPA authors have also done a very comprehensive job critiquing this work and rebutting these researchers' arguments. I will defer, however, the judgments of panel members more qualified than I in epidemiology, statistical design, molecular biology, and mutagenesis.

Dr. W. Michael Foster

Charge Question #7 – Appendix L

Question 7. *EPA solicited public comments on a July 2013 public comment draft of the IRIS carcinogenicity assessment of EtO and has revised the assessment to respond to the scientific issues raised in the comments. A summary of the major public comments and EPA's responses are provided in Appendix L. Has EPA adequately addressed the scientific issues raised in the public comments? For example, please comment on EPA's explanations for (i) its use of the lymphoid cancer grouping and (ii) combining unit risk estimates derived separately for the independent cancer types of lymphoid cancer and breast cancer to develop a total cancer unit risk estimate.*

Many of the public comments are related to cancer biology, and I do not have sufficient expertise in the area of cancer, and causal factors or their mechanism of action that are specific to lymphoid cancers. I have tried to review the EPA responses and placed consideration on the expressed rationale used by EPA in good faith, to address the public comments and following, are my summaries of the EPA responses and identified for each public comment.

<u>Public comments from ARASP, EOSA, and ACC.</u>

#1. EPA failed to comply with multiple guidelines, including Quality Act guidelines and 2011 NAS recommendations. Specifically, EPA failed to apply a transparent and systemic weight-of-evidence….

I interpreted the response by EPA to this 1st public comment to be certainly adequate and explanatory. In the response, attention was directed to Appendix J (Summary of major new studies since the literature cutoff date), and clearly stated that new studies noted by ACC in its public comments have been added to Appendix J. New studies included, Kiran et al, 2010, and Mikoczy et al, 2011, both of which were described and over-viewed by EPA in Section J2 of Appendix J. EPA in conclusion for the Kiran et al report, that this study added further support to the weight-of-evidence, but less than conclusive evidence of a causal association between EtO exposure and lymphohematopoietic cancers in humans (also quantitative risk estimates could not be derived). With respect to the Mikoczy report, EPA indicated that standardized and incidence ratios (SMRs and SIRs) for lymphohematopietic cancers in the Mikoczy study were consistent with an increase in lymphohematopoietic cancers; however EPA suggested that the study was underpowered for the analysis of lymphohematopoietic cancers and therefore the study contributed little to the weight of evidence for these cancers.

#2. Data quality evaluation should clearly describe the criteria used to deem a study as high quality.

A summary of the characteristics used by EPA in the EtO assessment was revised in order to more clearly respond to this public comment.

#3. Lymphohematopoietic and lymphoid cancers should not be grouped because they are derived from a different cell of origin.

EPA response, and based on prior support by the 2007 SAB, and seemingly quite defendable, was it is appropriate to combine lymphoid cancers, as the lymphoid cancer category is a grouping of cancers with a common lymphohematopoietic cell lineage.

#4. The evidence of breast cancer is too weak.

EPA response was to identify the following: that with no objection previously by the 2007 SAB, and with the understanding that the data base for breast cancer is limited (few studies with sufficient numbers of female breast cancer cases) more so than for the lymphohematopoietic cancers, and thus EPA position (and quite reasonable) was that based upon available evidence, breast cancer could be considered as a potential hazard from EtO exposure.

#5. EtO is a weak mutagen.

Essentially the EPA response was to agree that EtO is a relatively weak mutagen in comparison to anti-cancer agents and other reactive epoxides as identified in the report by Vogel and Nivard, 1998. EPA's agreement on mutagenicity was in part based on the results of the Vogel and Nivard report, and the comparison of EtO with 37 anti-cancer agents, which apparently are highly mutagenic by design.

#6. A mutagenic Mode of Action (MoA) is not supported by the most recent scientific evidence; other MoAs, specifically oxidative stress and cell proliferation, should be considered.

The EPA response was that a mutagenic MoA was operative in the carcinogenicity of EtO (this suggestion was supported by the 2007 SAB Panel) and that new evidence, including a mouse model of lung carcinogenesis, was not supportive due to several deficiencies in the mouse model (Parsons et al, 2013, and robustly overviewed in Appendix J). The EPA position on the Parsons report, was based in part upon a high degree of variability in the mutant fraction results from several of the dose groups (exposure was performed in the mice model by inhalation of EtO for several durations, i.e., 4, 8 or 12 weeks, to 0, 10, 50, 100, or 200 ppm) and the instability of the control results. With respect to regenerative cell proliferation as a result of EtO-induced cytotoxicity, as being a MoA, EPA response was essentially based on a lack of solid evidence that cell proliferation was a response and not solidly supported by published reports. With

respect to the Parsons report, it is not clear if in fact lung deposition fraction estimates for EtO in mice have actually been accomplished and which would impact suspected carcinogenetic dose effects attributable to EtO.

#7. EPA failed to incorporate the UCC data into the dose-response assessment. The NIOSH exposure assessment also suffered from limitations.

The clear impression given in the EPA response is that with a recommendation by the 2007 SAB Panel, the UCC data were not considered to be of sufficient quality to be a useful addition to the NIOSH study's data base
for the derivation of unit risk estimates. EPA qualified the use of the NIOSH regression model.

#8. Despite SAB recommendations, EPA used summary data rather than the individual data in the modelling of breast cancer mortality and lymphoid cancer.

The apparent objective of EPA's response, seems to attempt to clearly define their approach of using linear regression of the categorical data to develop the preferred unit risk mortality estimates for lymphoid cancer. In their approach, the highest exposure group was excluded. It is not clear to me why this was necessary. EPA stated that for breast cancer mortality data were also difficult to model and EPA also used a linear regression of the categorical data in their approach to develop the preferred risk estimates for breast cancer. It is also indicated by EPA that in the 2013 public comment draft, additional alternative models based on exposure data were developed and added to the assessment for comparison. These explanations of the EPA approach seem reasonable.

#9. EPA used a non-peer-reviewed supralinear spline model.

The response by EPA identifies the basis for their approach with a peer-reviewed publication and that the spline model will receive additional attention during the present SAB review.

#10. There are a number of modeling issues in addition to those mentioned in other comments, specifically flaws discussed in Valdez-Flores and Sielken (2013)……and over-predictions of the cancer deaths in the NIOSH study.

The response by EPA was to suggest that the referenced citations did not provide convincing evidence of flaws in the modeling; as well EPA added discussion to a section of Appendix J to specifically address concerns raised with respect to the Valdez-Flores and Sielken (2013) study.

#11. EPA should present both linear and non-linear extrapolation approaches.

The EPA response explains the background history with the 2007 SAB Panel and the EPA decision to not include a non-linear extrapolation approach and that the decision was clearly

explained in Appendix H. The concluding decision by EPA (not include non-linear extrapolation) was essentially defended by their being insufficient evidence at this point for a non-linear approach.

#12. Combining breast cancer and lymphoid cancer unit risk estimates is not justified and EPA did not discuss competing risks, different background populations, incidence vs. mortality, and the use of different exposure response models.

The EPA response clearly indicated the intent by EPA that its standard practice in IRIS assessments to estimate total cancer risk and not just the risk from individual cancer types, and that this practice is consistent with EPA guidelines and NRC recommendations. EPA also stated in the response, that the issue of different background populations (male and female) is now addressed in the assessment.

#13. EPA should reexamine its risk determination given background and endogenous levels of EtO; EPA's risk estimates are unrealistically high.

The EPA response identifies, using 2005 National-scale Air Toxics Assessment data, the average exposure concentration of EtO, and their own draft unit risk estimates (with adjustments for an assumed increased early-life susceptibility) to define the upper-bound estimates of the cancer risk resulting from a lifetime exposure to the average concentration of EtO (all sources). This seems very straightforward for estimating the risk of lymphoid cancer and breast cancer from predicted lifetime exposure to the average concentration above background.

#14. EPA should not derive occupational exposure limits for EtO.

The EPA response clearly identifies how other programs within EPA, i.e., Office of Pesticide Programs (OPP), has regulatory interest in occupational exposures that may result from sterilization uses of EtO. The response further suggests that the availability of IRIS unit risk exposure estimates permits OPP to review these estimates for its own risk assessments of occupational exposures (with the qualifier, that this has validity when the exposure response model is fairly linear over the relevant range of exposures). From a protection standpoint, this explanation by EPA seems reasonable.

Dr. Gary Ginsberg

Charge Question #1: Exposure-response modeling was conducted separately for lymphohematopoietic cancer mortality, with attention to lymphoid cancer, and breast cancer incidence and mortality. In the Cox proportional hazards models, a lag period was used to represent an interval before cancer death (or diagnosis, in the case of breast cancer incidence), or the end of follow-up, during which any exposure was disregarded because it was not considered relevant for the development of the cancer outcome observed. The lag period for each of the different cancer types was selected empirically based on statistical fit. These exposure lag periods were included in EPA's exposure-response analyses using other model forms for the derivation of cancer risk estimates. Please comment on whether the use of lagged exposure estimates in the derivation of cancer risk estimates and the selection of the lag periods used are clearly described and scientifically appropriate.

Response: Before addressing the issue of lag-fitting of the dose response model, I have the general observation that the animal-based potency value is approximately 10 fold less than the human-based value. While there are various potential biological reasons for this (e.g., differences in tumor types), it can be argued that animal studies have better opportunity to detect internal cancers such as mammary and lymphoid because every animal is fully examined for these outcomes; additionally there is no a priori reason to think that rodents would be less sensitive than healthy adult worker populations to a mutagenic carcinogen. Thus, it appears that the animal data do not support the human potency estimate. The draft document does not pay this much attention other than to say that the human data has primacy over rodent data due to no need for cross-species extrapolation. While this is a reasonable position to take, it will still be worthwhile for the draft document to recognize that the animal data are supportive of the findings of human mammary and lymphoid tumors due to tumor site concordance and then address what the potency difference may be based upon and how much uncertainty it creates.

One possible reason for the difference in potency may be related to the lag period-adjusted exposure estimate. By disregarding exposures that occur during the latency period, one is decreasing the cumulative exposure dose associated with the tumor outcome and thereby increasing the potency estimate. The biological basis for this exclusion is speculative as carcinogenesis requires numerous mutational events, some relatively late to complete the oncogenic process (malignant conversion, tumor progression). It may be that early exposure causes EtO-induced initiation that is sufficient to produce breast and lymphoid tumors in exposed workers and that the last 15-20 years of exposure would not be needed to manifest tumors as other genotoxic events would yield conversion and progression. However, it is plausible that there is a subset of breast and lymphoid tumors which would not have progressed without additional genotoxic insult caused by EtO through the complete time course of exposure. In fact it would be surprising if the additional years of EtO exposure did not yield additional risk.

Further, it is unclear whether someone who worked with EtO for 16 years only has the first year of exposure counted and the remaining 15 years disregarded. This would obviously have a huge effect on the dose response. Since lagging is not a routine practice for USEPA IRIS cancer derivations, this document should have been more transparent in the quantitative effect that lagging had on cancer potency estimate. It strikes me that the animal potency does not take into account dose lagging of exposure which is a potential reason for the potency difference across species. Given that the best dose response model fit included 15 year lagging, it is possible that this is the best approach. However, USEPA should present the model fit statistics and potency estimate for shorter and even 0 lagging periods.

Charge Question #5: Please comment of the accuracy, objectivity and transparency of the revised draft assessment, with particular emphasis on genotoxicity.

Response: The USEPA draft document does an excellent job of summarizing the EtO genotoxicity database accurately presenting the conclusion that it is a direct-acting agent as evidenced by the formation of DNA adducts and positive effects in a variety of in vitro and in vivo mutation and clastogenesis assays. This evidence is further strengthened by evidence of mutational and clastogenic effect in exposed workers. While the description of the database is adequate, the synthesis of this information into a mutagenic MOA could have been more systematic and complete. As currently constructed, USEPA asserts that EtO has a mutagenic MOA for all tumor types based upon: 1) EtO is a direct acting alkylating agent and these agents are typically able to cause mutation and cancer; 2) evidence with EtO of widespread genotoxic effects in vivo, in some cases at doses comparable to those causing tumors in animal studies, and in subchronic studies suggesting temporal association with cancer causation; 2) lack of alternative MOA concepts for EtO. There is also some discussion of mutagenic effect in key target tissues (e.g., Houle et al. evidence of mutation spectra shift in breast cancer tissue). However, Section 3.4 could be better organized around the evidence for a mutagenic MOA and thus lay the framework for defining mutagenic MOA in this document and for other chemicals going forward. My suggestion for the key elements of this framework are as follows, which is informed by the review by Eastmond (2012). <u>1) Characterization of the principal effects:</u> does the chemical interact with protein, DNA, redox cycles, cell cycle/rate of cell replication, apoptosis, signalling pathways, etc.; which are low dose and which are high dose effects; for EtO the primary effect is direct interation with DNA causing adducts; other effects occur including protein adducts and some evidence of oxidative stress; <u>2) Does the chemical produce mutagenic or clastogenic effects</u> – in what systems and along what dose response relative to adduct formation, cytotoxicity and cancer bioassay results; for EtO the literature is consistent in showing mutagenic and clastogenic effects across a wide range of test systems and at doses that are not cytotoxic; further, these genotoxic effects occur at doses below those required to induce tumors; it would be good to clarify whether specific DNA adducts are associated with these genotoxic effects but that may be a mechanistic uncertainty that does not negate the overall concept of mutagenic MOA for EtO but would detract from DNA adducts as a biomarker to

indicate dose response; 3) Does the chemical cause genotoxic effects in cancer target organs: these effects can include DNA adducts (weight increased if they are known to be promutagenic DNA adducts), mutational and clastogenic effects in the target organ; if such data do not exist, then are there any toxicokinetic (e.g., distributional factors, metabolic factors) or other mechanistic reasons to believe that genotoxic effects seen in vitro or in one cell type in vivo (e.g., PBL) wouldn't also occur in the target organ; for EtO there is evidence for mutational effects in breast tissue from mouse cancer bioassay (altered mutational spectra of tumor tissue – e.g., Houle et al. 2006) as well as altered mutational spectra in lung and other target tissue tumors (Hong et al. 2007); the fact that EtO-induced mutational spectra changes occur in tumor suppressor genes and oncogenes provides additional weight to a mutagenic MOA. Regarding lymphoid tumors, there is evidence for genotoxic effect of EtO in bone marrow and PBL from a number of studies. 4) Are there any non-linearities that would suggest that the mutagenic MOA does not continue to be operative at low or high dose levels? For EtO the DNA adduct dose response extends to very low doses, well below the cancer effect level. Toxicokinetic non-linearities have not been identified. 5) Temporal relationship: do DNA adducts and genotoxic effects precede the carcinogenic effect – for EtO, short-term and subchronic studies find evidence of genotoxic effects; 6) Alternative or contributing mechanisms: what other effects might explain or be major contributors to the oncogenic effect and how robust are these findings relative to the genotoxic findings; for EtO cytotoxicity, oxidative stress, alteration of signalling pathways may occur but there is no evidence that I am aware of that these effects occur along a dose response that would make them the primary effect at low dose. 7) Summarization of the cancer MOA via description of key events and whether they are consistent with a mutagenic MOA; for EtO these key events would appear to be: a) DNA adduct formation; b) mutation/clastogenesis; c) clonal expansion of altered cells; d) tumor formation. The statement of key events on Page 3-42 does a reasonable job of summarizing what is believed to occur with EtO but as I state above could be better supported.

A table which shows the dose response for DNA adduct and in vivo genotoxic effects in relation to doses shown to be carcinogenic in comparably dosed rodent studies would be helpful to establish relationships between genotoxity and carcinogenicity.

Charge Question #7 – Response to Public Comments, Appendix L

In general these responses were well reasoned and supported by the EtO database and standard risk assessment methods. For #4: Breast cancer evidence – USEPA can also use the evidence of mammary tumor increase in EtO-exposed mice to further support the use of this endpoint from the human study.

Reference: Eastmond, DA (2012) Factors influencing mutagenic mode of action determinations of regulatory and advisory agencies. Mutation Res 751: 49-63.

Dr. Steven Heeringa

CQ1: Exposure lagging. *Exposure-response modeling was conducted separately for lymphohematopoietic cancer mortality, with attention to lymphoid cancer, and breast cancer incidence and mortality. In the Cox proportional hazards models, a lag period was used to represent an interval before cancer death (or diagnosis, in the case of breast cancer incidence), or the end of follow-up, during which any exposure was disregarded because it was not considered relevant for the development of the cancer outcome observed. The lag period for each of the different cancer types was selected empirically based on statistical fit. These exposure lag periods were included in EPA's exposure-response analyses using other model forms for the derivation of cancer risk estimates. Please comment on whether the use of lagged exposure estimates in the derivation of cancer risk estimates and the selection of the lag periods used are clearly described and scientifically appropriate.*

For purposes of this review, the question of introducing a lag in cumulative exposures into the dose response modeling of risks for lymphoid and breast cancer mortality and incidence has three components: 1) is the introduction of the time censoring of exposure measures clearly described in the report?; 2) is the statistical methodology for choosing the length of the lag period appropriate?; and 3) is the introduction of the lag scientifically justified?. In this preliminary response, I will focus on 1 and 2 and leave 3 to panel members with more substantive expertise on the subject of latency period for onset of breast and lymphoid/lymphohematopoietic cancer and survival time to cancer mortality for some share of incident cases.

The draft IRIS report and its Appendices clearly describe the nature of the modeled latency for cancer incidence/death and the resultant time lag that is applied to cumulative exposures measures used in the preferred models of risk. Although the revised draft and Appendix D describe many new and varied trials at modeling dose response; the final selected models all retain the "optimal" exposure lag periods identified in the earlier published analyses of these data (Steenland et al. , 2003,2004): lymphoid cancer and lymphohematopoietic cancer- 15 years; breast cancer mortality-20 years; breast cancer incidence -15 years. In this earlier work, these selected lags were chosen based on goodness of statistical fit criteria for models of risk estimated for cases and controls in grouped exposure categories (as opposed to individual exposure data). Appendix D (pp. D24-D29) presents detailed output from a new trial of fitting Cox Regression (nested case control) models for incident breast cancer to cumulative exposure and log cumulative exposure with varying lag periods (lags of 0,5,10,15 and 20 years). Although I found no discussion of this trial or its evaluation in Appendix D, inspection of the results for the models that utilized the log of cumulative exposures show very little to distinguish the fit between the model that imposed a 15 year lag and that using no lag in cumulative exposure. This reanalysis appears to match the results of the original Steenland (2003) analysis which found nearly

equivalent quality of model fit for log cumulative exposure models with 0 or 15 year lag. (Additional covariates include date of birth (four cohorts?), parity and indicator of familial breast cancer). Given no strong statistical support for choosing one lag period over another, the draft report (P4-31, 32) concludes "The log cumulative exposure model with no lag was considered less biologically realistic than the corresponding model with a 15-year lag because some lag period would be expected for the development of breast cancer". To summarize, the position taken in the draft IRIS report appears to be that in the absence of strong statistical arguments in favor of 0,5,10,15, or 20 year lags for cumulative exposure, a lag period of 15 years (breast cancer incidence) is more biologically plausible and remains the choice for derivation of the risk models and the unit risk estimates. (Regarding the issue of the lack of statistically significant differences attributable to the choice of lag in the cumulative exposure (unit or log scale) measures, the total hazard in these Cox regression models of breast cancer incidence is heavily dominated by other covariates in the model including birth cohort, parity and family hx of breast cancer.)

In Appendix H (page H-8 and following) of the draft report, Professor Steenland responds to a recommendation of the 2007 SAB Panel, "The Panel encouraged the EPA to investigate potential instability that may result from the chosen time metric for the dose/response model and the treatment of time in the estimated exposure (e.g. log cumulative exposure with 15 year lag) that is the independent variable in that dose-response model". Despite the clumsy wording of this recommendation (which I am afraid I must take at least partial credit for), Dr. Steenland deduced the essence of the question and provided a comprehensive response on the issue of estimation of exposures prior to 1975 (in the absence of any sampling data prior to 1975). He also addresses the implication of the original exposure prediction model assumption (Hornung, et al, 1994) that calendar time effects (year) which were significant after 1978 were absent prior to 1978--allowing the predictions to pre-1978 exposures to be a function of the 1978 time effect (Figure 1 in Hornung et al.) and additive effects of other predictors in the model (exposure category, product type, product age, engineering controls, air volume of work area, etc.).

The modeling of risk from the NIOSH data set involves cancers that occurred during the period 1980-2004. Taking as an example the models of risk for breast cancer mortality and applying a 20 year lag to cumulative exposures implies that the last year of exposure eligible for inclusion in the cumulative value could have been as early as 1960 or as late as 1984 and that for most cases and controls included in the Cox model for matched cases and controls, the majority of the cumulative exposures would consist of estimates for a time period prior to the time ("after the late 70s.." Hornung et al., 1994) in which individual samples were collected and for which the prediction model relating exposure to job and job location characteristics was estimated. This remains an issue that I would like to explore in next week's SAB panel meeting.

The 2007 Panel raised the question of the potential for a "healthy worker effect" to dampen the dose response relationship at high cumulative levels of exposure. Appendix Section D.7 includes an interesting observation that "The degree to which the healthy worker survivor effect

confounds measured exposure-response trends is not known but it is likely that lagging exposure, as has been done here, diminishes such confounding". While this may in fact be a correct statement, the draft report should be careful to delineate that the choice of a lag in exposure should be supported by biological and statistical evidence and not simply adopted as a device to attenuate a potential source of bias in the underlying data.

Dr. Lawrence H. Lash

Response to Charge Questions:

CQ7. *EPA solicited public comments on a July 2013 public comment draft of the IRIS carcinogenicity assessment of EtO and has revised the assessment to respond to the scientific issues raised in the comments. A summary of the major public comments and EPA's responses are provided in Appendix L. Has EPA adequately addressed the scientific issues raised in the public comments? For example, please comment on EPA's explanations for (i) its use of the lymphoid cancer grouping and (ii) combining unit risk estimates derived separately for the independent cancer types of lymphoid cancer and breast cancer to develop a total cancer unit risk estimate.*

Response:

The EPA presents their responses to public comments received in July, 2013 as Appendix L of the new draft IRIS assessment. The section begins with a brief and clear summary of the comments received. The document lists the source of the comments and notes that excluding the three requests to extend the public comment period, a total of 16 comments were received. Because there was some repetition, some comments are grouped together, making 14 be the total number of unique comments to which the EPA responded.

Before assessing the responses of the EPA to each of the 14 specific comments, I would first like to make a general assessment of the nature of the comments received by the EPA, which primarily came from industry or industry organizations. In general, although several of the comments cite specific studies or make reference to the 2011 NRC report or the rules and regulations under which the EPA operates, the comments seem really designed to obfuscate and delay any new IRIS submission. Many of the comments are very loosely based on facts. Other comments are based on clear misinterpretations of established principles and longstanding practices. While some of these misinterpretations may be honest differences of opinion, many are clearly designed to place roadblocks in the way of getting this new IRIS document published.

Comment #1: This comment claimed that the EPA failed to follow NRC guidelines and failed to apply a systematic and weight-of-evidence approach. Comments like this seem designed to merely cast general doubt on the validity of the risk assessment and are not based on any solid facts. In fact, no evidence is provided that the EPA did not follow recommended guidelines. The EPA response is clear but could even be stronger. There are several places in the main document where the weight-of-evidence approach is discussed and justified. To strengthen the response to this question, some more detail listing places in the document where NRC and EPA guidelines

and the systematic and weight-of-evidence approach are explained and justified would be helpful.

Comment #2: The comment states that the EPA did not properly explain the criteria used to evaluate studies and deem them to be of high quality for inclusion in their analysis. Here either the people making the comment did not really read the study or are just making this comment to place a roadblock in the way. This is something that the document clearly explains in multiple places. Further, the document reiterates these criteria in a single summary in response to any concern that this point is not made clearly enough.

Comment #3: The comment states that lymphohematopoietic and lymphoid cancers should not be grouped because they are derived from different cells of origin. The response clearly states the rationale for grouping these together and notes that the 2007 SAB panel agreed with the logic of that grouping for comparison purposes. This response is clear and appropriate.

Comment #4: The comment states that the evidence for breast cancer is too weak. The response notes that the document acknowledges that the breast cancer database is more limited than that for other cancers. Further, the response notes that the previous SAB panel accepted the derivation of a unit risk factor based on that database. This response is clear and appropriate.

Comment #5: The comment notes that EtO is a weak mutagen. Both the response and the document itself never claim that EtO is a strong mutagen. The "weakness" of EtO as a mutagen as compared to many anti-cancer compounds and other reactive epoxides was clearly stated. In their response, the EPA provides further justification by noting that there is seldom a good correlation between mutagenic and carcinogenic potencies. This response is clear and appropriate.

Comment #6: The comment states that a mutagenic MoA is not supported by the most recent scientific evidence; other MoAs, specifically oxidative stress and cell proliferation, should be considered. There are two major issues here with regard to the MoA. First, the database concerning the MoA is rather complex, points that the document and the EPA response acknowledge. Second, and most significantly, the 2013 study cited in the comment is considered to be flawed and does not adequately argue that other MoAs besides direct mutagenesis are involved. The response clearly states that there is no clear support for the conclusions in Parsons et al. (2013). In the response, the EPA cites another recent study (Nagy et al., 2013) that does not support oxidative stress. The response provides a detailed discussion of the problems of inferring too much from K-ras mutation data. Even less data exist to support a proliferative MoA. The EPA response methodically presents the reasoning behind this conclusion.

Comment #7: The comment criticizes the EPA for failing to incorporate the Union Carbide Corporation (UCC) data into the dose-response assessment. It goes on to state that the NIOSH exposure assessment also suffered from limitations. The EPA response is concise and clear. This

issue is discussed in detail in the revised document and was supported by the 2007 SAB review. The NIOSH study meets the criteria of being a high-quality study much more strongly than the UCC data. This response is well-supported and appropriate.

Comment #8: This comment criticizes the EPA for using summary data rather than the individual data in the modeling of breast cancer mortality and lymphoid cancer despite SAB recommendations. Two key points are made in the response. First, the rationale for the modeling procedures used and their consistency with the previous recommendations of the SAB committee are noted. Second, the response notes that the current document adds additional models based on continuous exposure data and has added them to the assessment for comparison purposes. This response is appropriate.

Comment #9: A comment from two sources criticized the EPA for using a non-peer-reviewed supralinear spline model. The response notes that the model was published in 2011. Further, the response notes that use of the model will receive additional review by the SAB. This is another example of a comment meant to obfuscate as it was made in 2013, or two years after peer-reviewed publication of the model in question.

Comment #10: A comment was made regarding other concerns about the modeling procedures used and how they lead to over-prediction of cancer deaths in the NIOSH study. In response to concerns raised by the two publications cited in the comment, the EPA provided additional discussion in an appendix. Further, the EPA notes that the potential degree of over-prediction is far less than that claimed in the comment and the two papers. This response is appropriate.

Comment #11: A comment was made from three sources that the EPA should present both linear and nonlinear extrapolation approaches. This subject is discussed at great length in the main document and in an appendix. The response further notes that the previous SAB review agreed that there was no evidence to support use of a nonlinear extrapolation approach. This response is appropriate.

Comment #12: A comment was made from two sources that combining breast cancer and lymphoid cancer unit risk estimates is not justified, and EPA did not discuss competing risks, different background populations, incidence vs. mortality, and the use of different exposure-response models. In their response, the EPA first notes that breast cancer and lymphoid cancers were first modeled separately and then combined. Rationale for combining these unit risk estimates is explained in detail in the document. Further, the subject of competing and background risks is also discussed in detail in the document. Finally, the response concludes by noting the distinction between cancer incidence and cancer status. This response is appropriate.

Comment #13: A comment was made from three sources that EPA should reexamine its risk determination given background and endogenous levels of EtO and that EPA's risk estimates are unrealistically high. The EPA response explains how background rates for the cancers of interest

have been taken into account in the risk determination. They also note that in one of the comments (i.e., from the ACC), an unrealistic exposure concentration was used in arguing their point. This response is appropriate.

Comment #14: Two sources commented that EPA should not be deriving occupational exposure limits for EtO. The EPA response makes two clarifications. First, the EPA's Office of Pesticide Programs (OPP) is indeed responsible for deriving occupational exposure limits. Second, and more importantly, the response notes that such a derivation was not conducted in the present risk determination. Rather, the response notes that with the models used for the EtO cancer data, however, the unit risk estimate is not appropriate for the full range of occupational exposure scenarios of interest to the OPP. For the purposes of the OPP, the assessment provides sample risk estimates for exposure scenarios of interest to OPP for its own risk assessment of sterilization uses of EtO. OPP is indeed expected to conduct their own risk assessment in the near future.

Overall Analysis of EPA Response to Public Comments in Appendix L: The responses provided by the EPA are focused, generally complete, and delivered in good faith. This, unfortunately, cannot be said of the Public Comments.

Dr. Maria Morandi

Charge Question 5B. *Please comment on the accuracy, objectivity, and transparency of the revised draft assessment, with particular emphasis on the following sections, which are either new or substantially revised since the 2007 external peer review:*

- *Appendix H (EPA's responses to the 2007 external review comments), in particular the responses to the comments on endogenous EtO (p. H-4), a nonlinear approach (p. H-13 to H-17), and the cancer hazard characterization (p. H-3).*

Response

Overall, EPA was highly responsive to the comments from the panel that reviewed the 2006 draft EtO assessment. As a result, the revised draft assessment is much improved. This was a particularly challenging undertaking because the panel did not reached consensus on some of the issues critical to the outcome of the assessment, as reflected in the SAB 2007 EtO review report. As a member of that panel, I found it to be one of the most contentious I have experienced in terms of the extensive discussions on key scientific questions important to the assessment without achieving a consensus opinion among panel members. EPA should be commended for addressing the comments from both the majority and minority opinions of the panel report.

EPA has expanded the evidence and discussion presented in relevant sections of the Hazard Identification chapter (Chapter 3) in response to comments on the designation of EtO as carcinogenic to humans (preponderance opinion of the prior review panel) or likely to be carcinogenic to humans (minority opinion) strengthening the discussion on the weight of the evidence for supporting the "Carcinogenic to Humans" classification based on human and animal data including precursor information. As EPA indicates in the document and their response, the human precursor data are quite limited but consistent with a mutagenic MOA. Sections describing adduct formation have been expanded and are now more clear and transparent.

An introductory paragraph summarizing the contents of Chapter 3 has been added in response to the panel's recommendation. Another recommendation from the prior panel was to add "a more inclusive summary figure and/or table at the beginning of Chapter 3.0". EPA did not address this comment specifically, perhaps because it was not sufficiently clear. This recommendation was meant to include a brief summary of the key findings of the Hazard Assessment at the beginning of the chapter in some form. This is consistent with the new format for IRIS assessments, which includes a grey box at the beginning of this chapter highlighting the main conclusions of the Hazard Identification. A similar addition should be considered for Chapter 4 of the revised draft assessment.

EPA has been highly responsive in stating clearly the criteria for judging strengths and weaknesses of the epidemiology studies, which are summarized in a general form at the beginning of 3.1 but also applied clearly (and repeatedly) in the justification for selection of the NIOSH cohort studies as key for derivation of unit risk.

The decision by EPA not to transfer material from Appendix A in toto to the main body of the document (recommended by the SAB 2007 report) is well justified. In addition, it is consistent with NRC (2011) recommendations that the main body of the assessment focus on the critical data, rationales, and analyses used to support the unit risk derivation and that, as much as feasibly possible, description of other studies or analyses be summarized in Appendices with appropriate cross-referencing in the main body of the assessment. If anything, the current document could benefit from transferring more materials to appendices, although it is acknowledged that the revised draft assessment does not conform completely to the new format.

EPA has clarified its designation of the unit risk estimate as “weak” in the prior draft assessment, and section 3.5.1 provides a good evaluation of the strength of the weight of the evidence in term of Hill’s criteria for causality.

In response to the prior review, EPA has added additional literature and greatly expanded the sections of the assessment that describe endogenous EtO metabolic production as well as EtO-DNA adduct formation from external and internal sources (including from endogenous ethylene). Section 3.3.2 of the draft assessment and Appendix C provide a transparent and critical description of the available data (including studies that were not available at the time of the 2006 Draft Assessment) and recognize its limitations, especially as it relates to the application of analytical techniques that can resolve and quantify the differential contribution of external and endogenous EtO to protein and DNA adducts. Based on the discussion presented in the assessment and considering the weight of the evidence from human and animal studies, EPA’s conclusion that while endogenous EtO is likely to contribute to measurable risk - even significantly more so at low external exposure levels – it is unlikely to overwhelm the effect from external exposure is well supported.

With regard to consideration of EtO metabolic formation from external exposure to ethylene, as recommended by a minority of the members of the review panel, EPA judged that it would not be useful based on the limitations of the studies suggested by the panel and, therefore, made no changes in the assessment. Nonetheless (an also in light of the analyses presented on pages H-15-H17), it appears that recognizing this source of metabolic EtO and briefly describing its relevance to the assessment would complete the description of sources of endogenous EtO and their relative importance in terms of adduct formation. This could be easily done in Appendix C with a brief description in Chapter 3 cross referenced to this Appendix.

EPA has been responsive in providing an expanded and more balanced discussion of human and animal studies of precursor events that support a mutagenic MOA.

In response to the panel's recommendation that, while agreeing that the NIOSH cohort study provides the most robust set of data for unit risk derivation, EPA should consider all epidemiological data more broadly with special focus on the Union Carbide workers studies. In response, EPA has expanded the discussion of epidemiology studies with a clear description of the limitations of each potentially useful set of data. The selection of the NIOSH cohort and the decision not to combine these data with the Union Carbide cohort is well and transparently justified in the revised draft assessment.

Concerns about the differential reliability of the cumulative exposure with 15 yr lag metric used in the assessment, Dr. Steenland provided a reasonable response as to why this would not alter the analysis appreciably. At any rate, it is also important to note that the exposure estimates likely to be of lower reliability (because there were no exposure measurement data that could be included in the exposure model prior to 1979) are also likely to be much higher than the more recent exposures, and therefore, play a less important role in the current derivation of the POD.

The prior panel expressed multiple concerns about the use of categorical instead of individual data and applying different models to estimate risk in different parts of the dose-response curve, the appropriateness of the assumption of linearity for low dose extrapolation, aggregation of lymphohematopoietic cancers together, and derivation of LH cancer risk based on males only. The revised assessment bases risk estimates on the lymphoid-type cancers (the LH aggregate analysis is provided for comparison only) and both genders combined. The preferred models are now single models, so the comment from the panel about use of different models for different parts of the dose-response curve was addressed. EPA was highly responsive in addressing concerns about the use of categorical data for POD derivation, including obtaining the individual data for the NIOSH cohort and subcontracting with Dr. Steenland, the principal investigator of the NIOSH studies, to perform multiple analyses with these data (including individual and categorical exposure estimates) using alternative modeling approaches. In addition, there was also an attempt to expand on the error analysis of the NIOSH cohort exposure estimation (this could not be accomplished because the data files used in that assessment were no longer available). Results from the extensive additional analysis are well described in the revised draft assessment, both in Chapter 4 and in Appendix D, together with the rationale for supporting the decisions by EPA in model selection. It is important to emphasize that Dr. Steenland's involvement in the additional analyses is a major strength of the revised draft not only because of its intimate familiarity with the NIOSH cohort studies but his expertise in exposure modeling and occupational epidemiology. Likewise, the analysis and rationale (including more expansion on EPA guidance) for using low dose linear extrapolation are compelling. While there can be disagreements with the final selection of models for derivation of unit risk (this is an area that will be discussed by the panel in the current review of the revised assessment), it is clear that EPA should be commended for this effort and commitment of resources to address the concerns expressed in the SAB 2007 report, and that the revised assessment is transparent on this issue.

Concerns about the suitability of life table methodology for determination of LEC01 have been addressed. EPA provides a convincing rationale (especially since alternative approaches are not available) for using the BEIR IV algorithm. The response to the request to present the range unit risk estimates for the upper and lower 95% confidence limits of the EC01 is also reasonable.

Concerns about the extent of description of the rational for applying ADAFs has been address by expanding the corresponding section in Chapter 4. EPA has also provided detailed responses on specific concerns about specific issues of the modeling approach. The responses are extensive and detailed, so in this sense EPA has not ignored any of the concerns expressed in the SAB 2007 report.

Dr. Victoria Persky

Accuracy, objectivity and transparency of Appendix J

Appendix J summarizes major new studies since the last cutoff date

Transparency: The logic and progression of the literature review are clearly outlined and supported. Criteria for exclusion such as exposure studies, methods studies and reviews are reasonable with specifics well documented in Table J-1. Detailed discussions of the Kiran and Mikoczy studies appear accurate and objective. The conclusion that the Kiran study supports the weight of evidence for associations with lymphohematopoetic cancers without individual data allowing for quantitative risk appears justified. The argument regarding the Mikoczy study that the smaller numbers of breast cancer and lack of precision on exposure renders this study less suitable than the NIOSH study for unit risk estimates is credible as well. A few minor suggestions for the review of these papers are:

Kiran paper:

1. Page J-5 line 11 remove the word ”and” from the parentheses so that it reads “unexposed, low medium and high exposures”
2. Page J-6 line 1 after the word “evaluated” at the beginning of the line add “although the association with medium/high frequency is of borderline significance”

Mikoczy paper:

1. Page J-11 Table 2 third column – the derivation of the value 0.22 ppm x years for the 75th percentile isn’t clear. Presumably it is from the heading for Table 2 in the paper with the highest exposure group listed at > 0.22 ppm-years. Perhaps that should be noted somewhere in the review.

Discussion of the Valdez-Flores and Sielken 2013 paper and the Parsons 2013 paper are thoughtful. The Valdez-Flores article argues against the use of categorical data derived from individual data for determination of relative risks and that linear models should not assume an intercept fixed at 1 since it doesn’t take into account of baseline exposures. EPA in turn argues that especially, as is the case with ethylene oxide, when there is strong supralinearity a log-linear model exaggerates effects of small numbers of people with very high exposures. Similarly, estimation of the intercept rather than assuming a fixed intercept would inappropriately raise the base and overestimate rates in the unexposed. EPA also argues that exclusion of extremes and categorization mitigate effects of small numbers at very high levels and allows for more precise characterization of low and high does effects. These are thoughtful and reasonable responses to the Valdez arguments.

The discussion of the Parsons article is similarly thoughtful. This article examined mutations of K-ras codon 12 GGT to GTT, TAT and GAT in lung samples in a series of mouse experiments with exposure to doses of ethylene oxide varying from 0-200 ppm over 4-12 weeks. They found that there were preferential mutations to GAT (a mutation more often seen in spontaneous mouse lung tumors) with effects stronger earlier at lower doses and argue for early expansion of K-ras clone followed by death of the mutant cells. The authors infer that these findings support expansion of previous mutations rather than initial genotoxicity. EPA argues that the instability of the data, examination of only one codon, and lack of consistency with previous studies suggest that the data is unreliable and limits inferences that can be drawn. These arguments are reasonable and address issues raised by the article in a thoughtful fashion.

We were also asked to review previous articles by Morgan (1981) and Ambroise (2005) which were not discussed in detail. These were studies with very small numbers compared with those reviewed in detail. According to the October 22, 2014 memo they will be included in the final assessment.

Dr. Kenneth Ramos

Question 5a: *Accuracy, Objectivity, Transparency, with emphasis on 3.3.3 and Appendix C*

The revised draft is an update of the 2007 review and includes published studies for the period of ------. The summary tables included in the report should be revised to include key details of the cancer studies that are relevant to the interpretation of findings and to assign relative weights given to the evidence. Some of the weaknesses identified in this regard may originate from the inherent biases that may associated when evaluating new supportive evidence for an otherwise, well-accepted genotoxic and mutagenic chemical. The rationale for unit risk calculations within the context of MOA considerations and initial key events should be expanded to increase clarity of presentation. While the documents are excellent factual summaries of the published data, only limited insights were provided into the decision making process and relative weights assigned to the evidence. As suggested in some of the public comments provided, consideration should be given to differences in the biology for different tumor sites and the degree to which cell or tissue differences differentially impact MOA. For instance, reference to levels of different adducts was provided in page 3-29 without making a clear distinction regarding the putative or assigned biological significance for each of the different adducts. Additional experimental details about the separation of endogenous from exogenous adducts should be provided for the Marsden et al., 2009 paper.

Dr. Lianne A. Sheppard

Charge question 1: *Use of lags*

I agree that it is scientifically appropriate to lag exposure for cancer endpoints. I would prefer that the understanding of lag periods were developed from biological understanding. In general, basing lag choices on information other than the optimal fit in the same data used to base the risk assessment is preferable in order to not have the two sources of information so intertwined. I suggest that since such separate information is not available that a sensitivity analysis approach of presenting unit risk estimates for identical models for up to about 5 different lag choices (e.g. lags in the range of 10-25 years) be considered. The biological basis for the various lags should also be discussed.

Charge question 2: *Breast cancer incidence model selection*

2a: *Considerations of model selection and their application for breast cancer incidence*

The use of the two-piece linear spline model is appropriate. The selection of an identity link function rather than the more standard exponential link function for the Cox model is acceptable. The same model should be used for estimating both low-exposure cancer risks and risks from higher occupational exposures.

2b: *Reasonable models*

General comment applicable to all endpoints: In general, EPA seems to be reluctant to believe what the data are showing from a variety of adequately flexible models, namely that the risk rises very rapidly for low-dose exposures and then continues to rise much more slowly for all exposures above the lowest exposures. This suggests that the bulk of the harm from EtO comes from any exposure at all, and that once a small amount of exposure has been received, the additional risk from higher exposures is relatively less than the increased risk from just the initial small amount of exposure. The presentation of “reasonable” models considers model fit and some *a priori* (and not clearly articulated) notion about the acceptable shape of the dose-response function in the low-dose region. Since the data don’t conform to the *a priori* notion, EPA considers and in many cases selects “reasonable” models based on a linear regression of the categorical results, models that do a much poorer job reflecting the patterns in the data.

Consideration of reasonable models should address quality of the fit in the region of interest for risk assessment. Prioritizing sufficiently flexible exposure parameterizations (e.g. not linear) and

exposure functions with more local behavior (e.g. splines, linear and cubic) reduces the impact of highly exposed individuals on the risk estimates for lower exposures.

2c: *Knot identification approach*

This is an appropriate approach and transparently described.

Charge question 3: *Lymphoid cancer model selection*

General comments:

1. The models developed using individual-level continuous exposure data appear to be appropriate even though they are claimed to be unsuitable in the document. The cubic spline, two-piece linear splines, categorical, and log-exposure models all suggest that the risk rises rapidly with a small amount of exposure and then rises much more gradually for even higher exposures. I don't agree with EPA's conclusion that the linear regression of the categorical results is the preferred or even a good alternative to the other better fitting models applied to the individual-level exposure data.
2. In my opinion it is not acceptable to use one model for risk assessment for part of the exposure range and a different model for another part of the exposure range.

3a&b: *Comment on EPA's rationale for the linear regression of categorical results; comment on considerations for model selection, etc.*

See above general comments. I found EPA's rationale for their choice perplexing, easy to dispute, and not transparently communicated. There seems to be some *a priori* notion that using a rapidly increasing risk function for risk assessment is unacceptable even though all the evidence in the data suggest that there is a very steep dose-response in the low-dose range. That this general pattern is supported in multiple analyses and by both local and more global parameterizations of the exposure suggests to me that it should be trusted and used in the risk assessment.

For the occupational exposures, my comments for breast cancer pertain here.

3c: *Life table calculations*

These seem appropriate. The entire approach could be more transparently described for a broad audience. It would also be helpful to go through some more crudely estimated alternatives so readers can understand clearly all the different aspects of obtaining the unit risk and excess risk estimates without having to rely on the more complex life table analyses.

Charge question 4: *Uncertainty: Are the qualitative discussions clear, objective, and scientifically appropriate?*

Generally, yes. There could be more work done to document the uncertainty present that can be quantified. For instance, a full and clearly organized table of unit risk estimates for analyses that differ according to the various outcomes, subcohorts, link functions, exposure parameterizations, confounder adjustments, and SE estimation approaches (Wald vs. profile likelihood) would be very helpful to readers for getting a much better understanding of model uncertainty.

General comments on the document, focusing on Chapter 4.

1. In general the text is repetitive yet incomplete, difficult to follow, and omits some key information that will help readers better understand the data, analyses that have been done, and inferences drawn from the analyses. There is also lack of clarity in the presentation. Many statements are not well supported, yet some of these are repeated often. Nonstandard terminology is used for the various regression models, making the text more difficult to follow. The use of passive voice also contributes to text that is difficult to follow. Finally, the choice of tables and the information provided in them could be greatly improved.
2. Source of funding should be considered as a factor when reviewing and commenting on evidence presented in the peer-reviewed literature since it is well-known that conflicts of interest (e.g. due to source of funding) can influence the objectivity and disinterestedness of the authors, thus potentially tainting the findings and conclusions.

Specific Comments

1. P 4-1: Focus exclusively on the NIOSH cohort for human risk estimates is appropriate.
2. P 4-4: Considerations for selecting exposure-response models is appropriate. EPA could also add consistency of results when using more local vs more global exposure parameterizations (e.g. splines vs. exposure transformation or parametric form for the exposure).
3. Reliance on internal standard for estimating the RR rather than the external standard from the SMR analyses is appropriate. Given likely biases from healthy worker effects, the SMR analyses should not be given much weight. (The document is OK in this respect, but there are some references to the SMR results suggesting the data don't support an increased risk. This text should be refined.)
4. Using a nested case-control study with logistic regression approach to analysis as an approximation to the Cox model with time-dependent covariates is appropriate. It is a bit sloppy to then go on and discuss the results as though they come from a Cox model when the analysis is logistic regression and to switch back and forth between reference to OR's, RR's, relative rates, and hazard ratios.
5. P 4-5 line 21: The reference to Steenland finding a stronger dose-response relationship in males than females without additional context for that statement is misleading. The text

should be revised to indicate also that the dose-response patterns in males and females were generally consistent with each other. While the text as currently written may accurately reflect what Steenland reported, subgroup analyses can mislead and should not be relied upon unless they are prespecified and there is good biological reason that motivates focus on one subgroup over another (and this is determined before the analysis results are considered). Subgroups selected as a result of analysis are likely to reflect more random high bias than is desirable. From the evidence included in the report, there is no meaningful difference in EtO risk between genders. This concept is stated a few lines down and is appropriate. Reworking the entire paragraph so a sentence taken out of context won't mislead readers would be very helpful.

6. It is appropriate to combine genders for lymphohematopoietic cancers.
7. Table 4-2: Excluding ORs (RRs) by category mean/median from the table for the continuous models is not appropriate. The estimates and their 95% CIs at the category medians should be provided in the last column. (Note: I think it is more appropriate to summarize these categories by their medians rather than their means; ORs could be provided for category means as well.)
8. Considering both lymphoid and lymphohematopoietic cancers is reasonably justified.
9. Clarify the source of and appropriateness of the occupational vs. environmental conversion numbers. The text appears to imply that workers are exposed 240 days per year and breathe 10 m^3 of air per day. Is this what was intended? Is the reason that workers are assumed to breathe a lower amount of EtO per day than the general population to account for both higher ventilation rates when they are working and to address the fact that workers are only exposed for 8 hours per day?
10. P 4-8 l 13: The focus on the statistical significance of the males only result too easily allows the findings to be misinterpreted. Reword the text to reflect the findings more clearly by focusing on the general patterns and coefficients with their 95% CIs rather than on the presence or absence of statistical significance.
11. P 4-8 l 17: The entire discussion of the supralinear dose-response is problematic. This is the first instance but the problem is pervasive throughout the document. As is evident from multiple different approaches to parameterizing the dose-response, it is clear that for all the outcomes the risk increases steeply at low exposures and then at higher exposures it continues to increase but more slowly. One advantage of using the individual exposures is that one avoids the strong simplifying assumptions that the risk is constant within exposure categories, an assumption that is necessary with categorical exposures. It is not appropriate to state that "the results are unstable for low exposures" as though this is a problem. It is a feature of the supralinear pattern of the results that small changes in exposure at the low end of the exposure range correspond to large changes in risk. The concept of stability should not be equated with supralinearity, particularly since this equivalence is then used as a justification for discounting the compelling and generally consistent results from multiple differently parameterized models that allow a steep increase in risk at low exposures.
12. P 4-8 l 23-27: I don't understand the statement that a Cox model with cumulative exposure is "inherently sublinear". Is this a standard proportional hazards model with an exponential link

function and a linear exposure term? How is it inherently sublinear if the exposure term is linear? And how does this contrast with the Cox model with log-cumulative exposure as supralinear? The language used to describe the various models is confusing, nonstandard, and convoluted.

13. Figure 4-1 (and other related figures): Anchoring the Cox model fit with linear exposure term for 0 exposure at RR=1 is visually misleading. See comments on this point in Breslow and Day Statistical Methods in Cancer Research, Volume 1, p. 229. The statement at the bottom of the figure caption about the curves being comparable in shape alludes to this point; it still applies regardless of how the fits are presented visually.
14. P 4-10 first paragraph: I don't understand why low-exposure extrapolation is necessary in this dataset. The lowest quartile has a mean cumulative exposure of 447 and a median of 374 ppm-days. Is there a plan to extrapolate to cumulative exposures below this level? I agree with the 2007 SAB conclusions cited and the further analyses provided in this document strongly support their recommendation. I do not agree that it is preferable to do a weighted linear regression of the categorical results omitting the highest exposure category.
15. I agree that using two-piece linear splines is a good idea and was an appropriate choice for these data.
16. P 4-11 l 19-20: I don't agree that the below 100 ppm-days slope gives low confidence in this setting and that the small number of cases with low exposure leads to that conclusion. One important reason to use a regression model, particularly a flexible one like the 2-piece spline, is to "borrow strength" from all the data in order to get a better fit in this low exposure (or any other) region of the data. Similar comments apply to the 1600 ppm-days knot fit discussion.
17. P 4-11 l 31: I don't agree with EPA's determination. The spline models provide better information about the risk for low exposure. Furthermore, for more complete perspective, other exposure parameterizations, such as the log and square root of cumulative exposure model parameterizations should also be considered in the risk assessment in order to understand the sensitivity to model choice.
18. P 4-13, l 4: I think the median values are more appropriate than the mean values here because they better represent the category exposure levels. The categorical model is inherently assuming that all exposures in each category carry the same risk. Thus approximating the category exposure level by the median most closely approximates the exposures for all individuals included in that category (vs. the mean). The mean is a better summary of an individual's cumulative exposure when it is obtained from multiple measurements. The median is a better summary of the categories used in the regression model.
19. Table 4-4: I have trouble with some of the comments made in this table.
20. Table 4-5: Please provide the unit risk estimates where omitted. In other words, remove footnote *d*. These models are not unsuitable and should be presented for comparison.
21. P 4-17 l 11-12: I don't agree with the judgment that the low-exposure spline fit was excessively steep. This statement and similar ones (e.g. "these models were deemed unsuitable" on p 4-19 l 9) are inappropriate and inadequately supported.

22. P 4-17 l 12: I think a comment that a result that is 70% higher is "similar" needs justification. What range of relative estimates is considered "similar" in this context?
23. I agree that the assumptions made to convert lymphohematopoietic mortality to incidence are generally appropriate. The wording of the sentence starting on line 13 of p. 4-21 should be refined; with the clarifying information in parentheses I understood what was intended but the wording should be more transparent regardless.
24. P 4-23 l 30: The discussion of low exposure instability in the results is problematic. Similarly stating that parameterizing exposure linearly leads to stable results is also problematic.
25. P 4-26: There is a small amount of difference for the support of the 700 vs. 13,000 ppm-days knot models. See Tables D-2c and D-2f. These differences should not be magnified by the presentation in the text, as they currently are (lines 11-24).
26. P 4-28 l 16: I don't understand how the low dose extrapolation was needed in this dataset and how it was done. This comes up later as well, e.g. p 4-32 l 35. Please clarify.
27. P 4-28 l 25: Unit risk estimates for a variety of alternative models should be presented. Table 4-9 should be expanded with additional rows (for additional models considered in sensitivity analyses) and with additional estimates in the unit risk column.
28. P 4-30 l 9: I agree with restricting to workers there more than a year
29. P 4-30 l 15: It is appropriate to present results for the full cohort and the subcohort that could be located. It would be helpful to show some estimates from models with identical control variables for each cohort as currently the adjustment variables do not appear to be the same in the analyses for both cohorts.
30. P 4-30 l 23: I prefer a different approach to confounding control, namely *a priori* specification (to the degree possible) and some sensitivity analyses to determine how the results of interest change with introduction of different confounders.
31. P 4-30 l 26: I agree with combining in situ and invasive cases.
32. P 4-30: I agree that emphasis should be placed on lagged cumulative exposure rather than duration or unlagged exposures. I would prefer to see a set of sensitivity analyses w.r.t. lagging that consider a range of lags that can be biologically justified.
33. P 4-32 l 28-32: I have difficulty understanding statements such as this.
34. P 4-35 l 8-12: This is one reason to avoid the categorical exposure model results for risk assessment. It is also a reason to describe categories by their medians rather than their means.
35. P 4-35: I think the cubic spline and square root transformation models should be considered further in this document. Their unit risk estimates would provide valuable perspective for readers.
36. P 4-37 l 2-3: I don't agree that the low-dose risk function for log cumulative exposure is too steep and thus should have been omitted from further consideration.
37. P 4-38 l 16: Basic data descriptions should also be provided. It is very difficult to gain deep insight into the results without more descriptive information.
38. P 4-39; Table 4-12: It would be helpful to revise the comments in the table, particularly to focus on how models are distinguished w.r.t fit in the low exposure region.
39. P 4-41 l 13-14: Please include the square root and log cumulative exposure models as well. Similarly, omit the need for footnote *c* in Table 4-13.

40. P 4-44 l 1-4: I agree with this statement. It should be expanded to include justification for the selected model.
41. P 4-48 l 25-31: It is important that the authors convey clearly what is the observable range of exposures in these data. This has not been done. It is unclear to me that extrapolation is needed in this dataset. The second statement also patently ignores the risk pattern evident in the data. The authors need to justify how the estimates are highly dependent on the POD in this setting.
42. P 4-49 l 16-18: I don't agree since the linear model of categorical estimates tosses too much exposure information. This is the same concern voiced by the previous SAB panel.
43. P 4-49 l 30 ending phrase: agree. This should be documented with sensitivity analyses that show different unit risk estimates.
44. Much of the discussion on pp 4-48 l 20 to end of 4-49 is problematic.
45. P 4-53 l 14-17: I agree with this statement.
46. P 4-54 l 10-11: I think a comment that unit risk estimates that spanned less than a fourfold range gave similar results needs justification. What range of relative estimates is considered "similar" in this context?
47. P 4-57 l 14-16: I agree with this statement.
48. P 4-86 l 7-9: It appears that to do the conversion, there is an assumption that workers work 365 days in a year. Is this what was intended?
49. P 4-86 l 15-19: I don't agree that one model can be used for the risk assessment from environmental exposures and a different model for occupational exposures. Sufficiently flexible models have been fit that one model should be appropriate for both exposure levels. Evidence should be presented that the NIOSH cohort does not represent exposures that are consistent with environmental exposures of potential concern.
50. P 4-87 l 25: I disagree with the concluding statement of the sentence ending on this line.

Some confusing aspects of Chapter 4 that should be cleaned up

1. Inconsistency of reference to the parameter of interest as a relative rate, relative risk, odds ratio, and hazard ratio. In this setting they are all approximately the same. This should be stated clearly somewhere and there should be more consistency in the text.
2. There should be a table giving background disease rates for all cancers considered in this document. While the summary table won't include as refined estimates as the data included in the life tables used for estimation, the document authors need to empower readers to be able to conduct back of the envelope calculations. P 4-5 l 5: Finding the rates in the url given is not transparent.
3. Please comment why low-dose extrapolation is needed in these data or edit appropriately. (p 4-14, l 14) Similarly on p 4-15 the first full paragraph, how were the results of this study extrapolated? Aren't there exposures in the range of interest? Please clarify what is being said and what is being done here.

4. Please develop at least one unit risk estimate fully. Where possible, develop this using approximations (vs. the full life table analysis) so readers can understand each step of the risk estimation and so that readers are empowered to obtain approximate results using back of the envelope calculations. These can then be compared to the estimates obtained from the full life table analysis.
5. P 4-15 line 16: The last x (multiplication symbol) appears that it should be a division symbol. Also the same error appears on p 4-20 line 15.
6. P 4-20 l 19: Reference should be to footnote *e*.
7. Include unit risk estimates in the tables that are the final ones to be considered for policy. For instance, if $\mu g/m^3$ will be used for unit risk estimates, then a column should be added for this unit or the value should be presented in parentheses in the table. That way there is better correspondence between the tables and the text.
8. Tables 4-5 and 4-6: As part of the table captioning, tell readers why one estimate is shown in boldface font.
9. Table 4-7: Note that the units for the coefficient are not the same for the log cumulative exposure model as for the untransformed exposures. Also other tables, e.g., Table 4-10, 4-11.

Dr. Daniel Zelterman

Exposure lagging

I have never seen this before but it seems like a good idea.

It is explained well.

Comments:

Not clear whether the figure of "15 years" was estimated or not. Where does this number come from? Were other values tried?

Why Cox regression? (I suspect an answer is related to limitations of SAS. See below.) There are statistical methods specifically developed for multiply matched case-control studies. The proportional hazards method has low power relative to other methods.

Why are the case:control's matched at 100:1? This is excessive and gains very little additional power over a smaller number. Any ratio of 10:1 would be fine. Much more could be learned by identifying additional cases. Of course, these are hard to find. Little is gained in terms of power by identifying additional controls, which are much easier to find.

Table 4-2 is on the wrong scale. All those zeros look ridiculous. (What are exposure/days, anyway?) If necessary use 10-to-the-power to express such estimates and divide by the standard errors and express these as z-scores.

Overall: It seems that all computing is done in SAS. While SAS is great for managing the data, its statistical capabilities pale next to those in R. The letters SAS, like KFC, and AARP no longer stand for anything.

Question 3: Lymphoid Cancer model selection

I had a lot of trouble with Sections 4.1.1. Even putting aside what was done, it is not clear how this was motivated.

Again the Cox regression was used (WHY?) despite its low power. Again, is this motivated because the authors are forced to use SAS? Why is the "Extra risk" in (4-2) used as the outcome of interest? Is this because SAS estimates it? Were other outcomes considered? The risk estimates are not age-specific and this proves to be an awkward assumption.

Age-specific-rates are available from NCHS, specific to these cancers, not just all-cause-mortality. This was good. But how many people are diagnosed with lymphoid cancers in the US? A few? Thousands? Does it affect some ages more than others? Do the answers to these questions appear elsewhere in the report?

On page 4-8, "The only statistically significant Cox regression…" Does this mean that there was no exposure effect? Maybe the story should end there. Instead, I get the impression that the authors were not satisfied with this conclusion and needed to find some "$p<.05$." So, now we have Table 4-4, page 4-14 which lists a number of models that were tried.

This table of models will be confusing to a lay audience. ("Doesn't the EPA know what they are doing?" some will ask.) It is not clear what "lack of fit" means. What test determined this?

This section would have been so much easier to digest if it began with a simple statement about the model that WAS used and WHY it was used and WHAT is told us.

Aside: A Bayesian approach could be used to advantage. Use the age-adjusted rates from NCHS as a prior distribution on the unexposed rate. This would greatly enhance the estimation of low-dose exposure rates. Unfortunately, Bayesian statistics are not well received in many circles. (Sigh).

The spline approach is OK but not well motivated. Knots (the join points) are difficult to estimate because the likelihood is flat. The authors tried hard and found one. They explained this search well. But some smooth measure (quadratic? Loess or other smooth non-parametric?) would be much easier for us to explain to a non-technical audience. How do we motivate a model with a "corner" in biological terms? Even the regression on quartiles of exposure makes more sense. Quartiles are frequently used and well accepted in nutrition studies, where effects are small.

Section 4.1.1.3: Lifetime Extra Risk. This section would be much better if it included an estimate of how many workers are exposed to EtO and at what doses and for how long. (Even very rough guesses would be fine.) Estimate how many workers would develop lymphoid cancers if unexposed, on the basis of the NCHS data. Then estimate how many workers would develop cancers from exposure estimates obtained in this report. The estimated number 0.877 ppm is displayed but the reader has no way to put this in any perspective.