American® Chemistry Council

September 20, 2018

Information Quality Guidelines Staff
Mail Code 2811R
United States Environmental Protection Agency
1200 Pennsylvania Avenue, NW
Washington, DC 20460

Re: Request for Correction under the Information Quality Act: 2014 National Air Toxics Assessment (NATA)

Dear Sir or Madam:

The Ethylene Oxide Panel of the American Chemistry Council (ACC), hereby submits this Request for Correction under the Information Quality Act (IQA) of 2000, Section 515 of the 2001 Treasury and General Government Appropriations Act, Pub. L. No. 106-554, the Office of Management and Budget (OMB) Guidelines for Ensuring and Maximizing the Quality, Utility, and Integrity of Information Disseminated by Federal Agencies,[1] and the Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency (EPA).[2] ACC represents producers and users of ethylene oxide (EO).

ACC seeks the correction of EO information disseminated in the 2014 update to the National Air Toxics Assessment (NATA), released on August 22, 2018.[3] The 2014 NATA relies upon the "Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide (CASRN 75-21-8) In Support of Summary Information on the Integrated Risk Information System (IRIS)"[4] to determine the risk value for EO. As detailed below, the 2014 NATA does not meet the IQA's data quality requirements because the EO IRIS Assessment is not the best available science. Therefore, the 2014 NATA risk estimates for EO should be withdrawn and corrected to reflect scientifically-supportable risk values. Moreover, EPA should not use the EO IRIS Assessment's inhalation unit risk estimate (URE) of 5×10^{-3} per $\mu g/m^3$, which corresponds to a one-in-a-million increased cancer risk concentration of 0.1 parts per trillion (ppt), to calculate EO risk in its ongoing Clean Air Act (CAA) Section 112 risk and technology review (RTR) rulemakings and other regulatory actions.[5]

---

[1] 67 Fed. Reg. 8452 (Feb. 22, 2002) (OMB Guidelines).

[2] Available at https://www.epa.gov/sites/production/files/2017-03/documents/epa-info-quality-guidelines.pdf (EPA Guidelines).

[3] Available at https://www.epa.gov/national-air-toxics-assessment/2014-nata-assessment-results (2014 NATA).

[4] EPA/635/R-16/350Fa (December 2016) (EO IRIS Assessment).

As producers and users of EO, ACC members are directly impacted by the errors in the 2014 NATA. The risk estimates based on the EO IRIS value have significant regulatory implications for ACC member companies who produce commercial products of value to consumers using EO. Correcting these deficiencies will result in more accurate estimates of potential risk that will lead to improved regulatory outcomes, the dissemination of more accurate information to the public, and overall reduced misconception.

This Request for Correction is organized into four sections. The Executive Summary provides a high level overview of the key reasons why the 2014 NATA does not meet the objectivity, accuracy, integrity and utility requirements of the IQA and the OMB and EPA Guidelines due to its reliance on the EO IRIS Assessment. The second section provides background information on the 2014 NATA and the EO IRIS Assessment. The third section highlights the information in the EO IRIS Assessment that is not scientifically supportable. In the last section, each of the key deficiencies in the EO IRIS Assessment is discussed in detail with supporting scientific evidence.

I. Executive Summary

In the 2014 NATA, EPA relies on updated benchmarks for several substances, including EO. For EO, EPA updated its cancer risk calculations to reflect the URE in the EO IRIS Assessment. The use of the URE value, however, results in inaccurate and misleading conclusions about EO risk.

The EO IRIS Assessment is based on a supralinear spline slope for lymphoid and breast cancer exposure-response analyses from an epidemiology study conducted by the National Institute for Occupational Safety and Health (NIOSH). **This supralinear risk assessment model predicts high risk at low exposures, lower risk at higher exposures, and estimates an unrealistically low concentration of 0.1 ppt.** This 10^{-6} risk specific concentration (RSC) is the lower bound lifetime chronic exposure level of EO that corresponds to an increased cancer risk of one-in-a-million. **Both the supralinear slope and the RSC are implausible based on the epidemiological evidence and biological mode of action.**

---

[5] In a recently proposed RTR rule, EPA solicits comment on whether it should ban the use of EO for one of the source categories. *NESHAP; Surface Coating of Large Appliances; Printing, Coating, and Dying of Fabrics and Other Textiles; and Surface Coating of Metal Furniture Residual Risk and Technology Reviews,* 83 Fed. Reg. 46262, 46294 (Sept. 12, 2018).

In addition, these implausible levels lack utility for regulatory purposes. **The RSC in the EO IRIS Assessment is 19,000 times lower than the air-concentration equivalent yielding normal, endogenous levels of EO in the human body. Likewise, the RSC is orders of magnitude lower than ambient levels of EO.** Thus, if the EO IRIS Assessment is to be believed, normal human metabolism and/or breathing ambient air is sufficient to cause cancer. The EO IRIS Assessment does not provide a meaningful basis for assessing and managing risk for EO.

As outlined below, the EO IRIS Assessment is substantially flawed and can be corrected by using the approach published by Valdez-Flores et al. (2010),[6] which models potential mortality excesses for lymphohematopoietic tissue (LH) cancers from the two strongest epidemiological studies (NIOSH and Union Carbide Corporation (UCC)) using a log-linear Cox proportional hazard model. **Valdez-Flores et al. (2010) estimated ranges for the maximum likelihood estimate (MLE) and the 95% lower confidence limit of the environmental concentration corresponding to an extra risk of one in a million [LEC (1/million)] of, respectively, 1.5-9.2 parts per billion (ppb) and 0.5-1.2 ppb.** The major reason for the large difference between these values and the EO IRIS Assessment estimates is that the IRIS Program uses a supralinear spline model and Valdez-Flores et al. (2010) uses the log-linear Cox model.

EPA's cancer risk assessment guidelines caution that "a steep slope [i.e., supralinear] also indicates that errors in an exposure assessment can lead to large errors in estimating risk."[7] This is relevant to the EO IRIS Assessment because the NIOSH exposure model has a much higher level of uncertainty between the late 1930s and 1978 when there was inadequate (1976-78) or no exposure data (<1976) to independently validate the model. Furthermore, the NIOSH exposure model was modified when estimating exposures prior to 1978 by fixing the effect of a key variable (calendar year) in the model.

Specifically, Hornung et al. (1994) determined that Calendar Year is a major predictor of exposure in the model after 1978, but they did not allow this variable to impact exposures in the model prior to 1978.[8] Hornung et al. (1994) surmised that Calendar Year acts as a surrogate for improvement in work practices. **Thus, the arbitrary decision to alter the model prior to 1978 essentially assumes there were no evolving work practices in contract sterilizer facilities**

---

[6] Valdez-Flores C, Sielken RL Jr, Teta MJ. 2010. Quantitative cancer risk assessment based on NIOSH and UCC epidemiological data for workers exposed to ethylene oxide. Regul Toxicol Pharmacol, 56(3): 312-20.

[7] EPA, Guidelines for Carcinogen Risk Assessment (March 2005), at 3-19. Available at https://www.epa.gov/risk/guidelines-carcinogen-risk-assessment

[8] Hornung RW, Greife AL, Stayner LT, Steenland NK, Herrick RF, Elliott LJ, Ringenburg VL, Morawetz J. 1994. Statistical model for prediction of retrospective exposure to ethylene oxide in an occupational mortality study. Am J Ind Med, 25(6): 825-36.

between 1938 and 1977 that influence exposure to workers. The EO IRIS Assessment did not critically evaluate the assumptions and uncertainties of the NIOSH exposure model.

Moreover, the EO IRIS Assessment makes an unsubstantiated and counter-intuitive claim that the EO sterilization process was historically constant and stable prior to 1978. Yet, even the authors of the NIOSH study predict higher exposures before installation of engineering controls ("e.g., increased ventilation and better door seals") in 1978, when OSHA standards were higher.[9] **Below, we provide information on evolving regulatory standards, residue levels of EO, equipment, engineering and processing practices that indicate that the NIOSH exposure model incorrectly predicted that exposures would decrease in earlier years compared to the 1970s for the most exposed jobs (e.g. sterilizer operator). In general, *underestimating exposures will overestimate risk*, and the EPA cancer risk assessment guidelines caution that use of a supralinear model will further exacerbate the impact of these exposure errors.**

The rationale for selecting the supralinear spline model is based on incorrect statistical procedures and visual misrepresentation of the data. The EO IRIS Assessment incorrectly calculates the statistical significance (e.g., p- and AIC values) of the supralinear spline dose-response model because it fails to account for the statistical impact of the trial-and-error exploration of different arbitrary values used in the EO IRIS Assessment's dose-response model, such as the exposure level where the slope changes in the model from a very steep slope to a shallow slope (i.e. the "knot").[10] In addition, the figures used to compare visual fits use categorical data rather than the individual cases that were modeled. Once the individual cases are used, the log-linear Cox model fits the data just as well as the more complex and ill-advised supralinear spline model. The log-linear Cox model best meets the objective of selecting the more parsimonious model with fewer assumptions and variables.

Biologically, selection of the log-linear Cox model is more consistent with the mode of action for EO. This is supported by the EO IRIS Assessment, which concludes it is "highly plausible that the dose-response relationship over the endogenous range is sublinear … that is, that the slope of the dose-response relationship for risk per adduct would increase as the level of endogenous adducts increases."[11]

---

[9] Steenland K, Stayner L, Greife A, Halperin W, Hayes R, Hornung R, Nowlin S. 1991. Mortality among workers exposed to ethylene oxide. N Engl J Med, 324(20): 1402-07.

[10] *See, e.g.*, Li W, He C, Freudenberg J. 2011. A mathematical framework for examining whether a minimum number of chiasmata is required per metacentric chromosome or chromosome arm in human. Genomics, 97(3): 186-92.

[11] EO IRIS Assessment, at 4-95.

Both the UCC and NIOSH studies should be included in the dose-response modeling so that the risk estimates are based on the best available human data. Although the NIOSH study cohort is much larger, both studies have comparable power for males when considering the number of events of interest, i.e., lymphohematopoietic tissue cancers. The EO IRIS Assessment excludes the UCC cohort based primarily on a comparison of the exposure assessments for both studies. The EO IRIS Assessment dismisses the UCC exposure estimates as “crude,” “largely-uninformative,” “much less extensive,” and “greater likelihood for exposure misclassification,” as compared with the NIOSH study, which is described as “well-validated” and “high-quality.” These descriptions lack objectivity and obscure the fact that the majority of the UCC cohort exposure estimates are based on contemporary data from different plants with identical or comparable processes. **Although the NIOSH exposure model was validated with data after 1978, there were no contemporary data between the late 1930s and mid-1970s to validate the final model. Thus, the UCC exposure assessment uncertainties are no greater than the NIOSH study uncertainties and, therefore, are not a valid reason to exclude the UCC cohort.**

The EPA Science Advisory Board’s (SAB) peer review of the draft EO IRIS Assessment did not remedy the shortcomings of the final EO IRIS Assessment. The presumption of objectivity that sometimes attaches to documents that have been peer reviewed does not apply in this case because authors of the NIOSH study influenced the analysis of the data as well as the responses to the SAB’s comments. This influence compromised the objectivity and independent analysis of the NIOSH study, and especially the NIOSH exposure model, in the draft and final EO IRIS Assessments.

II. The 2014 NATA and the EO IRIS Assessment

The 2014 NATA uses emissions information to help state, local, and tribal air agencies identify which pollutants, emission sources, and places may warrant a better understanding for any possible risks to public health from air toxics. EPA further uses NATA results to improve data in emission inventories; identify where to expand air toxics monitoring; help target risk reduction activities; identify pollutants and source types of greatest concern; help decide what other data to collect; better understand risks from air toxics; and work with communities to design their own assessment.

The 2014 NATA results list EO emissions information across a range of categories, including location, cancer risks, hazard quotients, source type (e.g., stationary sources, mobile, airports, etc.). In building the NATA, EPA must select specific risk levels for certain air toxics that can lead to determinations of acceptable or unacceptable thresholds. Since air toxics have no universal, predefined risk levels that clearly represent acceptable or unacceptable thresholds, EPA makes case-specific determinations and general presumptions that apply to certain regulatory programs that further inform the interpretation of risk in the NATA. These benchmarks are drawn from a range of sources and updated. EPA notes that several substances' benchmarks were updated since the 2011 NATA, including EO. Specifically, EPA states that its risk value for EO was updated in 2016—the newly finalized IRIS value. As such, EPA updated its cancer risk calculations to reflect this new updated benchmark value. Due to the use of the EO IRIS value, more areas show elevated risks driven by EO in the 2014 NATA than in the 2011 NATA, even if emissions levels have stayed the same, or even decreased, in these areas.

The alleged elevated cancer risk driven by EO in the 2014 NATA has already caused alarm in some communities around facilities with EO emissions. This, in turn, has created media attention, and coverage of the issue has created further confusion and concern in the surrounding community. All of this could have been avoided had EPA relied on the best available science in calculating the unit risk estimate for cancer.

As discussed in detail below, the use of the updated EO IRIS value in the 2014 NATA and its Technical Support Document is extremely problematic given the EO IRIS Assessment's numerous shortcomings. A simple comparison of the results of the EO IRIS Assessment to the "real world," however, demonstrates its lack of credibility. Specifically, the RSC is 19,000 times *lower* than the normal, endogenous levels of EO in the human body. Likewise, the RSC is orders of magnitude *lower* than ambient levels of EO. Thus, if the EO IRIS Assessment is to be believed, normal human metabolism and/or breathing ambient air, without more, is sufficient to cause cancer. It strains scientific credibility to conclude that the EO IRIS Assessment presents a legitimate basis for determining risk for EO.

III. Request for Correction

The 2014 NATA relies upon the EO IRIS Assessment's inhalation URE of 5×10^{-3} per $\mu g/m^3$ to calculate EO risk. This URE implies a corresponding RSC of 0.1 ppt. The use of these values, however, results in inaccurate and misleading conclusions about EO risk because they are not supported by the scientific data. The RSC is also unrealistic, given that it is orders of magnitude lower than levels of EO in ambient air and levels that are consistent with normal, endogenous levels of EO present in human bodies.

A more reasonable and scientifically supportable approach to an exposure response analysis yields ranges for the MLE (1.5-9.2 ppb) and LEC (0.5-1.2 ppb) that are more than three orders of magnitude greater than the RSC.[12] Moreover, the ranges of MLE and LEC values are conservative because (a) extra risk was calculated despite no statistically significant slope in the exposure-response analyses; (b) the NIOSH data was included without adjustment for likelihood of underestimation of exposures; and (c) the limited evidence of cancer risk based on the entire body of epidemiologic evidence (see Appendix 2). The 2014 NATA risk estimates for EO should be withdrawn and corrected to reflect these risk values. Moreover, EPA should not use the EO IRIS Assessment's RSC of 0.1 ppt or URE of 5 x 10^{-3} per µg/m^3 to calculate EO risk in its ongoing CAA Section 112 risk and technology review or other rulemakings.

---

[12] Valdez-Flores et al. (2010).

A. The 2014 NATA Does Not Meet the Objectivity, Integrity, and Utility Requirements of the IQA and the OMB and EPA Guidelines.

Congress enacted the Information Quality Act (IQA) to "ensur[e,] and maximiz[e,] the quality, objectivity, utility and integrity of information (including statistical information) disseminated by Federal agencies" such as EPA.[13] The IQA required OMB to issue government-wide guidance, which each federal agency was to follow in its issuance of its own guidelines. The purpose of the EPA Guidelines is to apply the OMB Guidelines to the Agency's particular circumstances, and to "establish administrative mechanisms allowing affected persons to seek and obtain correction of information … disseminated by the agency that does not comply with the [OMB] guidelines…."[14] The 2014 NATA, therefore, must meet the OMB Guidelines as well as the EPA Guidelines.

OMB Guidelines include clear definitions to guide agency practices in adhering to the IQA. These include:

- "'Information' means any communication or representation of knowledge such as facts or data, in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms."[15]

- "'Influential,' when used in the phrase 'influential scientific, financial, or statistical information,' means that the agency can reasonably determine that dissemination of the information will have a clear and substantial impact on important public policies or important private sector decisions."[16]

---

[13] *See* Pub. L. No. 106-554. The IQA was developed as a supplement to the Paperwork Reduction Act, 44 U.S.C. §3501 et seq., which requires OMB, among other things, to "develop and oversee the implementation of policies, principles, standards, and guidelines to …apply to Federal agency dissemination of public information."

[14] Pub. L. No. 106-554.

[15] OMB Guidelines, at 8460.

[16] *Id.*

- "'Objectivity' involves two distinct elements, presentation and substance. 'Objectivity' includes whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner…. In addition 'Objectivity' involves a focus on ensuring accurate, reliable, and unbiased information. In a scientific, financial, or statistical context, the original and supporting data shall be generated, and the analytic results shall be developed, using sound statistical and research methods."[17]

- "'Utility' refers to the usefulness of the information to its intended users, including the public. In assessing the usefulness of information that the agency disseminates to the public, the agency needs to consider the uses of the information not only from the perspective of the agency but also from the perspective of the public. As a result, when transparency of information is relevant for assessing the information's usefulness from the public's perspective, the agency must take care to ensure that transparency has been addressed in its review of the information."[18]

The 2014 NATA is influential scientific risk assessment information and must adhere to a rigorous standard of quality.[19] The 2014 NATA is "influential" scientific risk assessment information as set forth in the EPA Guidelines because it "will have or does have a clear and substantial impact (i.e., potential change or effect) on important public policies or private sector decisions" and involves "controversial scientific … issues."[20] Results from the NATA are used by government agencies, non-governmental organizations, and air quality experts to gauge which hazardous air pollutants (HAP) and emission sources may raise health risks in certain places. These places are then given more attention and EPA uses the NATA to, among other things, target ways to achieve risk reduction.

The NATA can also lead to the development of local community-supported plans to reduce emissions as presented in each NATA version's results. Additionally, the National Research Council (NRC) has recognized the NATA as one of the largest EPA efforts to "develop baseline cancer risk estimates and hazard index calculations using dose-response information and exposure estimates."[21] In this context, NRC further acknowledges the importance of the NATA as a "tool for exploring control priorities" and its function "as a preliminary attempt to establish a baseline for tracking progress in reducing HAP emissions."[22] Therefore, the 2014 NATA, and its underlying data, must adhere to a rigorous standard of quality, including meeting the higher standard of reproducibility.

---

[17] *Id.* at 8459.

[18] *Id.*

[19] Quality includes objectivity, utility, and integrity.

[20] *See* EPA Guidelines, at 19-20 (internal citations omitted); OMB Guidelines, at 8455.

[21] National Research Council, "Air Quality Management In the United States" (2004), at 247. Available at https://www.nap.edu/read/10728/chapter/1.

With regard to the analysis of risks to human health, safety and the environment maintained or disseminated by the agencies, the OMB and EPA Guidelines also require either adoption or adaption to "the quality principles applied by Congress to risk information used and disseminated pursuant to the Safe Drinking Water Act Amendments of 1996 (42 U.S.C. 300g-1(b)(3)(A) & (B))."[23] In ensuring the objectivity of influential scientific risk information (i.e., the substance of the information is accurate, reliable and unbiased), the EPA Guidelines have adapted these principles by requiring the use of the "best available science and supporting studies" and the collection of data using by "accepted methods or the best available methods" using "a 'weight-of-evidence' approach that considers all relevant information and its quality."[24]

EPA has failed to apply a transparent and systematic weight-of-evidence approach in assessing the cancer risks of EO exposures in the 2014 NATA. Moreover, as detailed below, because the 2014 NATA relies upon the EO IRIS Assessment to determine the risk value for EO, the 2014 NATA is not based on the best available science.

B. The EO IRIS Assessment Does Not Meet Scientific Standards from Multiple Standpoints.

The EO IRIS Assessment is not the best available science because it: (1) exclusively relies on a NIOSH study despite its flawed exposure assessment; and (2) applies a supra-linear spline model, which is implausible based on the epidemiological and biological evidence and deficient due to statistical miscalculations and visual misrepresentations.

1. The EO IRIS Assessment incorrectly describes the NIOSH exposure model as a "state-of-the-art" validated regression model to estimate historical exposures prior to 1978. In fact, this "state-of-the-art" validated model was tested with post-1978 data only and arbitrarily altered for years prior to 1978. Specifically, a variable considered to be a major predictor of exposure after 1978 was not allowed in the model to impact exposures prior to 1978. The

---

[22] *Id.*

[23] *See* EPA Guidelines, at 22-23; OMB Guidelines, at 8460.

[24] *See* EPA Guidelines, at 21-22. "In this approach, a well-developed, peer-reviewed study would generally be accorded greater weight than information from a less well-developed study that had not been peer-reviewed, but both studies would be considered." *Id.* at 26. The definition of best available science mirrors that articulated in Chlorine Chemistry Council v. EPA, 206 F.3d 1286 (D.C. Cir. 2000), referring to "the availability at the time an assessment is made." *See* EPA Guidelines, at 23.

reliability, validation and likelihood of exposure misclassification prior to 1978 were not objectively evaluated.

2. The results of NIOSH's statistical model for exposures prior to 1978 were not provided in the 2014 Draft EO IRIS Assessment or in the cited NIOSH publications. In the appendices of the final EO IRIS Assessment, two new figures (Figures D-22 and D-33) present new information on estimated exposures by worker, but no explanation or critical evaluation was added. There is a lack of transparency in the EO IRIS Assessment of these influential data used to derive the EO cancer slope factor.

3. The EO IRIS Assessment repeatedly asserts that the NIOSH exposure estimates were well-validated using a state-of-the-art model, when in fact there was no validation of exposure estimates prior to 1978. These assertions regarding verification procedures are incorrect for the late 1930s to 1978.

4. In response to public and SAB comments questioning the lower than expected exposures in earlier years predicted by the statistical regression model, the IRIS Program states that the decrease is related to the sterilizer volume. In other words, the model predicts that smaller sterilizer volume results in lower exposures. This response essentially uses the output of the model to answer a question about whether the model assumptions are correct, instead of independently verifying the validity of these assumptions. This circular reasoning does not address the underlying concern of whether the model assumption that Sterilizer Volume has an inverted parabolic (that is, an upside-down U-shaped) relationship with predicted EO exposure is correct. It also does not address whether other factors that might result in increased exposure during early years were properly accounted for in the model.

5. The EO IRIS Assessment makes the unsubstantiated claim that "the sterilization processes used by the NIOSH cohort workers were fairly constant historically, unlike chemical production processes, which likely involved much higher and more variable exposure levels in the past."[25] In fact, there was an evolution in technology and practices associated with the sterilization processes between the late 1930s and early 1970s. Data and information from industrial sterilization operators and the literature refute this claim.

6. Comparisons of relative reliability made between the NIOSH and UCC studies are inaccurate. These comparisons were a key basis upon which the IRIS Program rejected the UCC Study as a source of epidemiology study data for cancer risk assessment. The EO IRIS Assessment does not acknowledge and appropriately consider limitations of the NIOSH

---

[25] EO IRIS Assessment, at 4-4.

exposure assessment posed by low extrapolations of NIOSH cohort exposures to EO prior to the late 1970s without any corroborating data or any supporting engineering/process considerations derived from or directly relevant to that period of time.

7. The EO IRIS Assessment relies solely on the NIOSH study of sterilant workers and fails to incorporate the important findings from the UCC study of workers in EO producing and using operations. The IRIS Program considered and characterized three factors in its selection of the NIOSH study: cohort size, exposure data, and confounding. Based on these factors, the IRIS Program dismissed the UCC study as a basis for EO cancer risk estimation. In considering cohort size, the IRIS Program ignored the most important comparison—the number of lymphohematopoietic tissue cancers, not the total cohort size.

8. The use of the supralinear spline model for the lymphoid and breast cancers in the final EO IRIS Assessment is based on an invalid statistical analysis. Because the analysis did not correctly calculate degrees of freedom associated with that fitted model, it contains erroneous measures of absolute and relative goodness of fit of that model. When both the p-values and Akaike Information Criterion (AIC) values characterizing fit quality are corrected, the supralinear spline model does not fit the NIOSH lymphoid tumor data statistically significantly better than the log-linear Cox model.

9. The selection of the supralinear spline model for the lymphoid tumors is also based on misleading illustrations of "visual fits" that do not convey either the actual data that were fit or the relative goodness of fit to these data of log-linear and supralinear spline models. Only in a footnote does the IRIS Program indicate that the visual comparison misrepresents the log-linear model being compared. Consequently, and erroneously, the fit to the data appears far worse than the supralinear spline model. The data plotted in that figure also were summary data that misrepresent the true magnitude of the scatter of the data that were used for model fitting.

10. The selection of a spline model as the preferred model for EO cancer risk estimation assumes a supralinear increase in tumor response in the low-dose exposure region with a subsequent plateauing of response at higher exposures. The body of cancer epidemiologic studies, including the NIOSH studies, does not support such a pattern of risk. While certain NIOSH sub-analyses suggest increases in male lymphoid tumors and female breast cancers, the findings are limited to the highest cumulative exposure groups, not the lowest.

11. The use of a supralinear spline model for cancer risk estimation is inconsistent with the assumed mode-of-action of EO toxicity and tumorigenicity. Such a model predicts higher risk at low exposures compared to risks predicted at higher exposures, which is contradicted by the well-understood mode of action of EO in experimental animals and humans

as described in the EO IRIS Assessment. Thus, the EO IRIS Assessment relies on human cancer risk estimates based on spline-model dose-response extrapolations that are internally inconsistent with its own evaluation of the mode of action of EO. The mean air concentration equivalent to the endogenous concentration in non-smoking humans with no known EO exposures is 1.9 ppb (range 0.13-6.9 ppb; continuous), which is 19,000 times greater than the EO IRIS RSC of 0.1 ppt.[26] An alternative LEC (1/million) of 0.5-1.2 ppb is a more pragmatic, science-based approach for EO risk assessment.

12. The statistical, epidemiological and biological evidence does not support the selection of supralinear spline models to fit the NIOSH study data in the EO IRIS Assessment. A more scientifically sound conservative alternative is to use the Valdez-Flores et al. (2010) approach, which incorporates all the available data from the two strongest human studies (NIOSH and UCC). This approach has been adopted by the European Commission's Scientific Committee on Occupational Exposure Limits.[27]

IV. Because the 2014 NATA Relies Upon the EO IRIS Assessment to Determine the Risk Value for EO, the 2014 NATA Is Not Based on the Best Available Science.

1. The EO IRIS Assessment incorrectly describes the NIOSH exposure model as a "state-of-the-art" validated regression model to estimate historical exposures prior to 1978. In fact, this "state-of-the-art" validated model was tested with post-1978 data only and arbitrarily altered for years prior to 1978. Specifically, a variable considered to be a major predictor of exposure after 1978 was not allowed in the model to impact exposures prior to 1978. The reliability, validation and likelihood of exposure misclassification prior to 1978 were not objectively evaluated.

The EO IRIS Assessment's evaluation of the cancer potency of EO is dependent on an analysis of commercial sterilization worker exposure conducted by NIOSH. The NIOSH EO data for the sterilization work cohort were nearly all collected between 1978 and 1986 at 20 different facilities, but included just seven mean values based on 23 exposure measurements for the period 1976-77.[28] Ultimately, of the 20 facilities, 16 facilities were eliminated from the exposure assessment for lack of personal sampling, documentation of sampling, or links of sampling to job categories.

---

[26] Kirman CR, Hays SM. 2017. Derivation of endogenous equivalent values to support risk assessment and risk management decisions for an endogenous carcinogen: Ethylene oxide. Regul Toxicol Pharmacol, 91: 165-72.

[27] *See* Recommendation from the Scientific Committee on Occupational Exposure Limits for ethylene oxide, SCOEL/SUM/160 (June 2012).

[28] Hornung et al. (1994).

Based on the available worker data, the workers included in the NIOSH study cohort were employed in the sterilization industry as early as the 1930s. Noting that "there were no measurement data prior to 1976," Hornung et al. (1994) describe the statistical model[29] developed to estimate NIOSH EO-cohort worker exposures based on data collected after 1978. That model was applied to estimate worker exposures over a large timespan (1935-1975) during which not a single observed measurement was available to validate the application of that model extrapolation procedure.

Although the NIOSH statistical regression model estimated exposure measurements after 1977 with reasonable reliability, Hornung et al. (1994) highlighted that post-1978 regulatory standards and consequent progressively stringent operational EO-exposure controls accounted for the pronounced decreasing trend in measured NIOSH-cohort EO exposures that occurred after 1978. Prior to 1978, these EO standards and controls were largely or entirely absent. Thus, they were irrelevant to most of the 1935-1975 timespan, during which time the NIOSH statistical model was applied to estimate historical worker exposures without any empirical physical-modeling basis for direct validation.

The final statistical model selected to predict the natural logarithm (ln) of EO exposure included two nonlinearly modeled variables which were determined to be the two most EO-predictive variables identified: Calendar Year ("Year") and Sterilizer Volume ("Cubic Feet"). These two variables were each modeled to have an inverted parabolic relationship to predicted ln(EO) levels, resulting in predicted peak EO exposures to occur during 1978 as a function of Year. Hornung et al. (1994) note that their final statistical model arbitrarily set the value of Year to be 1978 for all years prior to 1978, explaining that:

> Since we felt that the decrease in ETO levels after 1978 (independent of engineering controls) was explained by improved work practices after ETO was identified as a potential carcinogen, we set each predicted ETO level prior to 1978 equal to the predicted level in 1978. Variation in exposure levels prior to 1978 were modeled as a function of the remaining terms in the model with the calendar year effect fixed at 1978. Therefore, there was no extrapolation by calendar year prior to 1978.

---

[29] Steenland NK, Stayner LC, Griefe AL. 1987. Assessing the feasibility of retrospective cohort studies. Am J Ind Med, 12: 419-30; Greife AL, Hornung RW, Stayner LG, Steenland KN. 1988. Development of a model for use in estimating exposure to ethylene oxide in a retrospective cohort mortality study. Scand J Work Environ Health, 14(Suppl 1): 29-30.

Thus, the “validated” model was arbitrarily and selectively altered for years prior to 1978 by fixing the calendar year value to 1978. Nonetheless, for the same period prior to 1978, the model still predicts that lower EO sterilizer volumes were associated with lower occupational EO exposures—a prediction made without any independent, pre-1978 measurement-based or physical-modeling-based evidence supporting such an association during that period. The IRIS Program should have questioned the reliability and validation of the model prior to 1978, and objectively considered the likelihood of exposure misclassification during this period.

2. The results of NIOSH’s statistical model for exposures prior to 1978 were not provided in the 2014 Draft EO IRIS Assessment or in the cited NIOSH publications. In the appendices of the final EO IRIS Assessment, two new figures (Figures D-22 and D-33) present new information on estimated exposures by worker, but no explanation or critical evaluation was added. There is a lack of transparency in the EO IRIS Assessment of these influential data used to derive the EO cancer slope factor.

A basic standard quality expectation for a peer-reviewed publication of a statistical model for exposure is that the results section should include summary of the output of the model; in other words, the estimated exposures resulting from the model. Neither the NIOSH exposure modeling publications nor the NIOSH epidemiology studies that rely on this model provide any descriptive summary of exposures estimated by the model prior to the late 1970s. The IRIS Program should have independently evaluated the exposure data, especially after ACC provided the summary of NIOSH exposures by job (reprinted below as Figure 1).

Figures D22 and D23 in the EO IRIS Assessment are graphs of estimated annual exposures for the entire cohort by worker, but not by job. However, there is no discussion or analysis of these graphs in either Appendix D or the main report. These figures are less informative in understanding how the NIOSH exposure model estimated exposure by job because these figures are based on each worker who could have different job assignments. Nevertheless, the 95th percentile of annual exposures of the NIOSH cases in Figure D-23 has a very similar pattern of exposures as the job with the maximum exposure in Figure 1 below.

As described below, neither Hornung et al. (1994) nor the IRIS Program offer any realistic explanation for the counterintuitive trend backward in time from the late 1970s that is predicted by the NIOSH statistical regression model, other than such a trend just happens to be what that statistical model predicts. Thus, there is a lack of transparency and independent critical evaluation of the exposure estimates of the NIOSH exposure model in the EO IRIS Assessment.

Moreover, the derivation of the NIOSH statistical regression model can no longer be reproduced, because the raw data on which it was based no longer exist.[30]

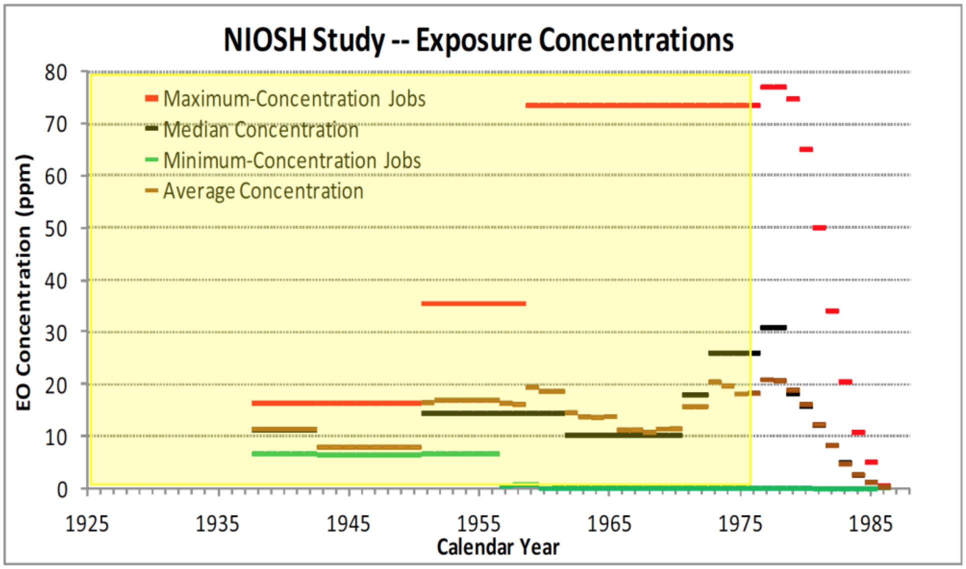
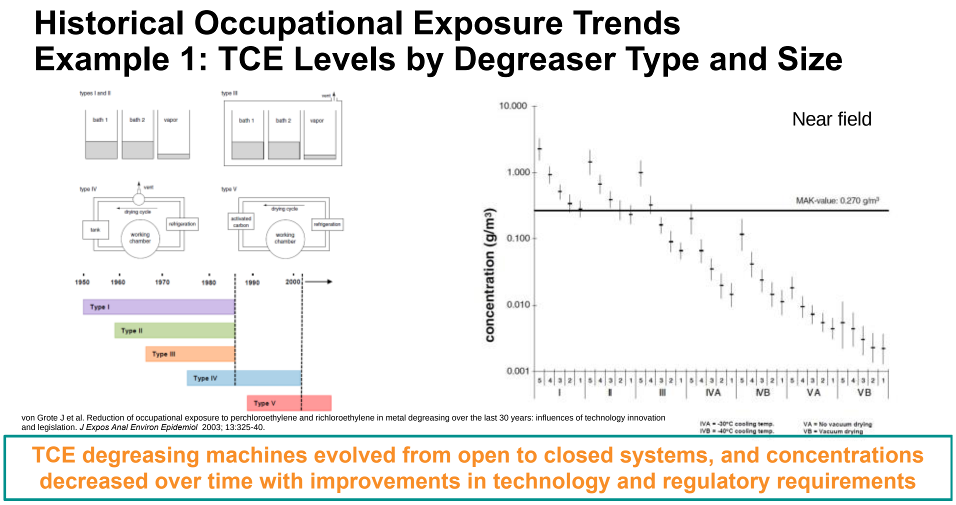


Figure 1. NIOSH statistical regression model predictions of 8-hour time-weighted average exposure to EO by job in each calendar year. This summary data for each job was provided by NIOSH and was used to estimate exposures for participants in the NIOSH cohort based on job code. This figure appeared on page 173 of Appendix M ("Comments on NIOSH Exposure Papers: Greife et al. (1988) and Hornung et al. (1994)") of Comments on the Revised External Review Draft Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide, Docket ID No. EPA-HQ-ORD-2006-0756 submitted to EPA by ACC on October 11, 2013, but did not appear either in Hornung et al. (1994) or any of the draft EO IRIS Assessments reviewed by SAB.

---

[30] Appendix H (Summary of 2007 External Peer Review and Public Comments And Disposition) of the EO IRIS Assessment states, "[i]n response to the panel's suggestion that the Hornung analysis represents an 'invaluable opportunity' for further analysis of the impact of possible errors in exposure estimation, the EPA investigated the possible use of the 'errors in variables' approach (page 27 of the panel report). Steenland visited the NIOSH offices in Cincinnati in order to review the data and assess whether it would support an errors-in-variables analysis. Unfortunately, the electronic data files used in the [NIOSH] exposure analysis were no longer available, so that analysis based on the errors-in-variables approach was not possible." *Id.* at H-28. Thus, the raw data on which NIOSH relied to derive its statistical regression model used to extrapolate historical NIOSH-cohort exposures to EO prior to the late-1970s, when measures of workplace EO first began to be made, no longer exist—implying that there is no longer any way to validate the claim by Hornung et al. (1994) that their model was able to predict the 85% of the variation in log values of EO concentrations measured starting in the late-1970s. Even if that claim were true, it has no logical bearing on the ability of that model to generate accurate extrapolations of occupational exposure to EO back in time prior to the late 1970s when, as emphasized by Hornung et al. (1994), occupational conditions were quite different because none or virtually none of many sterilization technology changes and sterilization workplace practices, which only began to be adopted starting in the late 1970s to greatly reduce EO exposures (as reflected by NIOSH-cohort exposure measures made starting in the late 1970s to which the NIOSH statistical regression model was fit), were in place prior to that time.

The pattern shown in Figure 1 indicates generally lower exposures for earlier time periods when the crudest technology was used under the least stringent worker protection standards. The SAB considered this pattern to be "surprising," as discussed in greater detail in Section 4, below. Indeed, the pattern of the NIOSH exposure data by job in Figure 1 is the reverse of patterns of historical exposure levels from published studies of exposures to volatile chemicals through time with improvements in technology and increased worker protection requirements[31] as illustrated in two relevant examples (Figures 2 and 3).



Historical Occupational Exposure Trends
Example 1: TCE Levels by Degreaser Type and Size

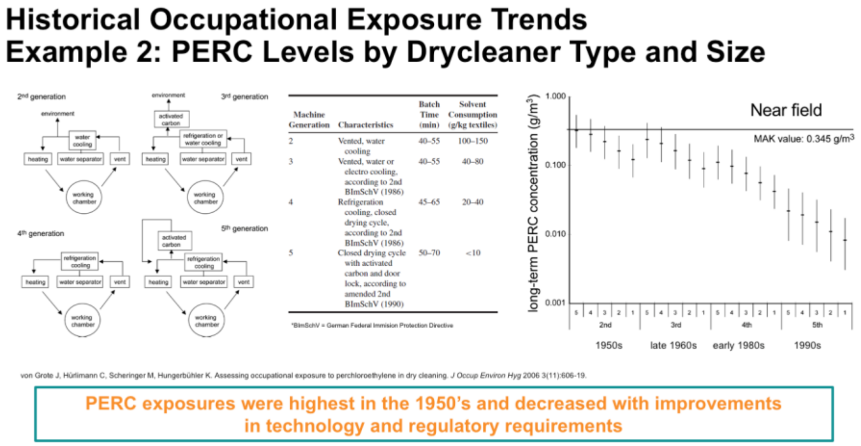
von Grote J et al. Reduction of occupational exposure to perchloroethylene and richloroethylene in metal degreasing over the last 30 years: influences of technology innovation and legislation. *J Expos Anal Environ Epidemiol* 2003; 13:325-40.

TCE degreasing machines evolved from open to closed systems, and concentrations decreased over time with improvements in technology and regulatory requirements

Figure 2. Historical occupational exposure trends, Example 1: TCE levels by degreaser type and size. Source: von Grote et al. (2003b).

---

[31] *E.g.*, von Grote JHM. 2003a. Occupational Exposure Assessment in Metal Degreasing and Dry Cleaning – Influences of Technology Innovation and Legislation. Doctoral Dissertation, Swiss Federal Institute of Technology, Zürich. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.628.1123&rep=rep1&type=pdf; von Grote J, Hürlimann C, Scheringer M, Hungerbühler K. 2003b. Reduction of occupational exposure to perchloroethylene and trichloroethylene in metal degreasing over the last 30 years: influences of technology innovation and legislation. J Expo Anal Environ Epidemiol, 13: 325-40; von Grote J, Hürlimann C, Scheringer M, Hunger K. 2006. Assessing occupational exposure to perchloroethylene in dry cleaning. J Occup Envir Hyg, 3: 606-19.



Historical Occupational Exposure Trends
Example 2: PERC Levels by Drycleaner Type and Size

Machine Generation	Characteristics	Batch Time (min)	Solvent Consumption (g/kg textiles)
2	Vented, water cooling	40–55	100–150
3	Vented, water or electro cooling, according to 2nd BImSchV (1986)	40–55	40–80
4	Refrigeration cooling, closed drying cycle, according to 2nd BImSchV (1986)	45–65	20–40
5	Closed drying cycle with activated carbon and door lock, according to amended 2nd BImSchV (1990)	50–70	<10

*BImSchV = German Federal Immision Protection Directive

von Grote J, Hürlimann C, Scheringer M, Hungerbühler K. Assessing occupational exposure to perchloroethylene in dry cleaning. *J Occup Environ Hyg* 2006 3(11):606-19.

PERC exposures were highest in the 1950's and decreased with improvements in technology and regulatory requirements

2nd generation
environment
water cooling
heating
water separator
vent
working chamber
3rd generation
environment
activated carbon
refrigeration or water cooling
heating
water separator
vent
working chamber
4th generation
refrigeration cooling
heating
water separator
vent
working chamber
5th generation
activated carbon
refrigeration cooling
heating
water separator
vent
working chamber
Near field
MAK value: 0.345 g/m³
long-term PERC concentration (g/m³)
1.000
0.100
0.010
0.001
5 4 3 2 1 5 4 3 2 1 5 4 3 2 1 5 4 3 2 1
2nd
3rd
4th
5th
1950s
late 1960s
early 1980s
1990s

Figure 3. Historical occupational exposure trends, Example 2: PERC levels by dry cleaner type and size. Source: von Grote et al. (2006).

3. The EO IRIS Assessment repeatedly asserts that the NIOSH exposure estimates were well-validated using a state-of-the art model, when in fact there was no validation of exposure estimates prior to 1978. These assertions regarding verification procedures are incorrect for the late 1930s to 1978.

Assertions made in the EO IRIS Assessment about independent evaluation of model estimates are inaccurate. Table 1 lists the statements in the EO IRIS Assessment related to the UCC and NIOSH exposure assessments.

Table 1: List of EO IRIS Assessment statements regarding UCC or NIOSH exposure assessment

Page Number	Description of UCC exposure	Description of NIOSH exposure
1-1		Had a well-defined exposure assessment for individuals
1-2		"high-quality" study based on several attributes, including availability of individual worker exposure estimates from a high-quality exposure assessment
1-4		Retrospective exposure estimation is an inevitable source of uncertainty in this type of epidemiology study; however, the NIOSH investigators put extensive effort into addressing this issue by developing a state-of-the-art regression model to estimate unknown historical exposure levels using

Page Number	Description of UCC exposure	Description of NIOSH exposure
		variables, such as sterilizer size, for which historical data were available.
3-5	Crude exposure assessment, with a high potential for exposure misclassification	
3-6		… the exposure model and verification procedures are described in Greife et al. (1988) and Hornung et al. (1994). Briefly, a regression model was developed to allow estimation of exposure levels for time periods, facilities, and operations for which industrial hygiene data were unavailable. The data for the model consisted of 2,700 individual time-weighted exposure values for workers' personal breathing zones, acquired from 18 facilities between 1976 and 1985. The data were divided into two sets, one for developing the regression model and the second for testing it. Seven out of 23 independent variables tested for inclusion in the regression model were found to be significant predictors of EtO exposure and were included in the final model. This model predicted 85% of the variation in average EtO exposure levels.
3-7		Good-quality estimates of individual exposure
3-8	"cruder" especially for highest exposure	Based on a validated regression model
4-3 and 4-4	Exposure assessment is much less extensive than that used for the NIOSH cohort, with greater likelihood for exposure misclassification, especially in the earlier time periods when no measurements were available (1925-1973). Exposure estimation for the individual workers was based on a relatively crude exposure matrix that cross-classified three levels of exposure intensity with four time periods. The exposure estimates for 1974-1988 were based on measurements from air sampling at the West Virginia plants since 1976. The exposure	This is in contrast to the NIOSH exposure assessment in which exposure estimates were based on extensive sampling data and regression modeling. In addition, the sterilization processes used by the NIOSH cohort workers were fairly constant historically, unlike chemical production processes, which likely involved much higher and more variable exposure levels in the past.

Page Number	Description of UCC exposure	Description of NIOSH exposure
	estimates for 1957-1973 were based on measurements in a similar plant in Texas. The exposure estimates for 1940-1956 were based loosely on a "rough" estimate reported for chlorohydrin-based EtO production in a Swedish facility in the 1940s (Hogstedt et al., 1979). The exposure estimates for 1925-1939 were further conjectures based on the Swedish 1940s estimate. Thus, for the two earliest time periods (19251939 and 19401956) at least, the exposure estimates are highly uncertain. (See Section A.2.20 of Appendix A for a more detailed discussion of the exposure assessment for the UCC cohort.)	
4-5		It was judged to be substantially superior to the UCC study with respect to a number of key considerations in particular, in order of importance: (1) quality of the exposure estimates …
4-60	largely uninformative in terms of assessing the unit risk estimates derived from the NIOSH study because of the crude exposure assessment used in the UCC study	

The EO IRIS Assessment does not critically evaluate the uncertainties of the NIOSH linear regression model, and does not clarify that the NIOSH model was not validated with any data prior to 1978. In the appendices, similar deficiencies pertain to assertions concerning measures applied purportedly to validate the NIOSH statistical regression model,[32] purported empirical and unbiased bases for the NIOSH statistical regression model,[33] and purportedly unlikely inaccurate characterization of exposure by the NIOSH statistical regression model and its purported validation despite nonexistence of original data upon which it was derived.[34]

NIOSH historical extrapolations of occupational EO exposures prior to the late-1970s, were, as described by Hornung et al. (1994), "derived from a regression model based on observed measurements." This regression model was applied to extrapolate worker exposures over a large timespan (1935-1975), during which not a single observed measurement was available to validate the application of that extrapolation procedure, and only a small subset of measures was available during 1976-77. Although the NIOSH statistical regression model reliably estimated exposure measurements made after 1977, Hornung et al. (1994) highlighted that post-1977 regulatory standards and consequent progressively stringent operational EO-exposure controls accounted for the pronounced decreasing trend in measured NIOSH-cohort EO exposures that occurred starting in 1978. Prior to 1978, EO standards and controls were largely or entirely absent. Thus, they were irrelevant to most of the 1935-1975 timespan.

---

[32] *See* EO IRIS Assessment, Appendix A, at A-14.

[33] *See id.*, Appendix D, at D-75.

[34] *See id.*, Appendix H, at H-27 – H-28.

4. **In response to public and SAB comments questioning the lower than expected exposures in earlier years predicted by the statistical regression model, the IRIS Program states that the decrease is related to the sterilizer volume. In other words, the model predicts that smaller sterilizer volume results in lower exposures. This response essentially uses the output of the model to answer a question about whether the model assumptions are correct, instead of independently verifying the validity of these assumptions. This circular reasoning does not address the underlying concern of whether the model assumption that Sterilizer Volume has an inverted parabolic (that is, an upside-down U-shaped) relationship with predicted EO exposure is correct. It also does not address whether other factors that might result in increased exposure during early years were properly accounted for in the model.**

During the review of the 2014 draft EO IRIS Assessment, the SAB questioned the general pattern of historical exposures that were lower in some or all years prior to 1975. The SAB had specifically requested EPA to address this issue in a substantive manner (i.e., using historical, physicochemical, and/or engineering facts or models independent of the NIOSH statistical regression model itself). The SAB noted:

> The SAB is also concerned that public commenters had exposure data from the NIOSH cohort that the EPA did not have. For instance, a few selected graphs were presented in **public comments to the Augmented CAAC that indicated exposure predictions for** four jobs in two of the fourteen plants **showed lower exposures in some or all years prior to 1975. The SAB was provided only a few carefully selected examples, and thus was unable to assess the extent of these surprising data.** This is an uncertainty that can easily be ruled out. **Upon reviewing the model equation in Hornung et al. (1994), the SAB finds the surprising historical behavior to be unlikely** and could be explained by changes

> in processes in specific plants, rather than some failure of the model to capture historically larger exposures. The EPA should ensure that they obtain all relevant data released from NIOSH to members of the public.[35]

Figure 1 above shows that the "surprising historical behavior" characterized by the SAB as "unlikely" does not pertain only to a few specific jobs in different plants, but is a general pattern going back in time prior to the late-1970s. EPA's response to the SAB's concern was:

> contrary to public comments made at the SAB meeting, the NIOSH EtO exposure patterns are not anomalous, but rather reflect the underlying changes in variables predicting exposure over time. One of the principal drivers of the NIOSH exposure levels was the cubic feet of the sterilizers used [see Table III, Hornung et al. (1994)]. It was not uncommon in these plants for sterilizer volume to have increased over time as the demand for EtO-sterilized products increased. **Increased sterilizer volume generally resulted in higher predicted average exposures until the late 1970s, when increased controls were used after it became known that EtO might be dangerous**.[36]

The IRIS Program provided quantitative examples illustrating the point emphasized in the quote above for two different plants, in effect illustrating that the response is consistent with the NIOSH statistical regression model defined in Tables III and VI of Hornung et al. (1994). However, the response is circular and, thus, nonresponsive to the SAB concern, because it relies on the same statistical regression model to attempt to validate its assertion that "increased sterilizer volume generally resulted in higher predicted average exposures until the late 1970s."

The NIOSH regression model predicts that EO exposure levels are proportional to an inverted parabolic (upside-down U-shaped) function of sterilizer volume. This function reaches a maximum predicted EO exposure level at a sterilizer volume value of approximately 4,000 ft^3. This regression function is estimated entirely from measurement data obtained nearly exclusively after 1977. However, NIOSH does not explain a plausible physical basis for this complex exposure/volume relationship observed nearly exclusively after 1977. Although this relationship explains a statistically significant amount of variation in the available EO measures, NIOSH offers no convincing evidence that such a relationship must also reliably apply to periods prior to 1978. Hornung et al. (1994) point out that regulatory constraints, sterilization operation, and sterilization technology all differed greatly from prior to 1978 vs. in/after 1978; they emphasize that in 1978, efforts to control EO exposure began to be implemented on an accelerated basis.

---

[35] Science Advisory Board Review of the EPA's Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide (Revised External Review Draft - August 2014) (Aug. 7, 2015), EPA-SAB-15-012 (2015 SAB Review), at 18 (emphasis added).

[36] EO IRIS Assessment, Appendix I, at I-26 – I-27 (emphasis added).

None of the three methods applied by Hornung et al. (1994) to validate their statistical regression model[37] is capable of providing any direct form of validation or verification of historical EO exposures actually incurred by the NIOSH cohort. The NIOSH regression model makes that prediction, based on its statistical regression fit to historical EO measurements that only began in the late 1970s, without any other empirical, physical-modeling, or engineering rationale upon which to establish even the plausibility of that model prediction (e.g., based on independent published literature, historical data, physical/compartmental modeling, or any type of reasoning whatsoever bearing on whether sterilizer chamber volume per se is or is not expected to have correlated with or determined historical EO exposure levels prior to the late-1970s).

Hornung et al. (1994) note that pounds of EO used each year served as a surrogate measure of potential EO exposure, but that since such EO utilization data "were not available for all plants in the study, the size of the sterilizer units (in cubic feet of capacity) was substituted after we determined that there was a high degree of correlation between these two variables." However, in order to achieve sterilization efficacy, EO concentrations used in sterilization chambers have remained approximately constant over time—*regardless* of the volume of sterilization chambers used—except insofar as EO concentrations used are well known (and were reported by experienced EO industry workers in interviews discussed below) to have *increased* going backwards in time from the late 1970s, because higher concentrations of EO were used in earlier decades during the evolution of sterilization operations and technology.

Likewise, because utilization of internal sterilization chamber volume has remained fairly constant over time, independent of reduced chamber volume going back in time from the late 1970s, opening of each chamber door and storage of off-gassing sterilized materials resulted in similar immediate concentrations of EO exposure to nearby workers. Reduced chamber volumes going back in time implied that greater numbers of such smaller chambers had to be used to process approximately the same load of sterilized material per plant. To the extent that smaller amounts of sterilized material were processed by plants earlier in time, then those processes are certain to have occurred in smaller facilities, implying that going back in time since the late-1970s there was either an increase (as noted above) or no substantial change in the mass-of-EO-used to workspace-volume ratio that determined the time-weighted average EO concentration to which sterilization workers were exposed throughout that period (particularly for the most heavily exposed workers).

---

[37] Hornung et al. (1994) explain that, in the absence of historical exposure data to perform such verification, they applied a three-phase evaluation procedure consisting of 1) a statistical cross-validation procedure applied to a subset of post-1978 empirical measures of EO, 2) comparison of predictions made a by "a panel of 11 industrial hygienists familiar with ethylene oxide levels in the sterilization industry" to the latter subset of empirical data gathered subsequent to 1978, and 3) an evaluation of the ability of the statistical model to explain the empirical variance exhibited by the entire set of empirical measures of (as noted above, nearly all post-1977) EO exposures available for the NIOSH cohort.

Of greater significance, EPA's response does not take into account critical variables, such as level of EO residue in sterilized materials based on the number of air washes used, the length of time sterilized materials were stored prior to return to customers, and where they were stored relative to chamber operations—variables that changed substantially over the decades of EO sterilization prior to the late 1970s. Historical (pre-late-1970s) estimates of NIOSH cohort EO exposure rely on historical extrapolations made only by the NIOSH statistical regression model that were driven primarily by a correlation primarily between chamber volume and post-late-1970s measures of EO exposure. Operational changes that could have influenced EO exposure concentrations prior to 1976/78 were not investigated.

Even the NIOSH study expected higher historical exposures that would be influenced by the absence of engineering and regulatory controls: "Exposure levels are likely to have been higher [than "the late 1970s"], however, before the installation of engineering controls, when the OSHA standard was 50 ppm instead of the present 1 ppm."[38] Moreover, in the 1940s and 1950s, the MAC-TWA and TLV-TWA were 100 ppm.[39] In 1978, the U.S. Food and Drug Administration (FDA) published proposed "maximum residue limits" of 5-250 ppm for medical devices for human use that are sterilized with EO. Prior to 1978, there were no regulatory standards to reduce residues on medical devices, so the residues were around 10–30,000 ppm depending on the type of material.[40] But the IRIS Program failed to take this information into account when modeling the data.

5. The EO IRIS Assessment makes the unsubstantiated claim that "the sterilization processes used by the NIOSH cohort workers were fairly constant historically, unlike chemical production processes, which likely involved much higher and more variable exposure levels in the past."[41] In

---

[38] Steenland K, Stayner L, Greife A, Halperin W, Hayes R, Hornung R, Nowlin S. 1991. Mortality among workers exposed to ethylene oxide. N Engl J Med, 324(20): 1402-07, at 1406.

[39] ACGIH. 2001. Ethylene Oxide: TLV® Chemical Substances 7th Edition Documentation.

[40] Ernst RR and Whitbourne JE. 1971. Toxic residuals. In the Study of the requirements, preliminary concepts, and feasibility of a new system to process medical/surgical supplies in the field, pp. 46-57, Appendix pp. 1-2, Contract No. DADA17-70-C-0072. U.S. Army Medical R&D Command, Washington, D.C. (Defense Documentation Center Accession No. AD890320 and AD890321).

[41] EO IRIS Assessment, at 4-4.

fact, there was an evolution in technology and practices associated with the sterilization processes between the late 1930s and early 1970s. Data and information from industrial sterilization operators and the literature refute this claim.

Interviews conducted by Exponent, Inc. with three former sterilization operators who began work in the mid-1960s and early to mid-1970s (one was a member of the NIOSH cohort) confirmed operational differences in the sterilization operations in the 1960s and 1970s, and in earlier decades, relative to operations post-1978. This new interview information is supported by information and data in the technical literature on sterilization operations in early decades, including high EO residue levels in and rates of EO off-gassing from EO-sterilized medical materials,[42] and by current quantitative measures of in-chamber EO concentration during sterilization operations after single and multiple air washes that were transmitted to Exponent, Inc. by an industrial sterilization company. These data indicate that the EO IRIS Assessment's assumption that the sterilization processes were fairly constant between the late 1930s and early 1970s is incorrect.

These data also indicate that the variables in the NIOSH model that predicted exposures after the mid-1970s do not capture important potential sources of exposures to sterilizer operators prior to the 1970s:

a. Technology improvements for worker protection such as back venting and use of aeration processing rooms to degas sterilized materials were implemented post 1978. Thus, the presence or absence of back venting or ventilated aeration rooms may help discriminate exposures after 1978, but not between the late 1930s and 1977.

b. Pre-1978 commercial sterilization operations typically included at most only a single post sterilization air wash (relative to numerous washes used typically in later decades); in a current sterilization unit using 100% EO, an EO concentration of 17,200 ppm was measured in chamber air after a single wash cycle. Fewer wash cycles result in much higher peak exposures when opening the chamber doors, as well as higher residue levels remaining on the pallets of sterilized material. These higher residue levels contribute to higher exposure levels to those working in areas where pallets are stored.

---

[42] Perkins JJ. 1969. Principles and Methods of Sterilization in Health Sciences, 2nd ed. Charles C. Thomas, Springfield, IL; Bruch CW. 1972. Toxicity of ethylene oxide residues. In: Phillips GB, Miller WS, eds. Industrial sterilization, Duke University Press, Durham, NC, at 119-23; Bruch CW. 1981. Ethylene Oxide sterilization—technology and regulation. Industrial ethylene oxide sterilization of medical devices: process design, validation, routine sterilization, AAMI Technological Assessment. Report No. 1-81. Arlington, VA: Association for the Advancement of Medical Instrumentation, at 3-5; Roberts RB, Rendell-Baker L. 1972. Aeration after ethylene oxide sterilisation. Failure of repeated vacuum cycles to influence aeration time after ethylene oxide sterilisation. Anesthesiol, 27(3): 278-82; Stetson JB, Whitbourne JE, Eastman C. 1976. Ethylene oxide degassing of rubber and plastic materials. Anesthesiol, 44(2): 174-80; White JD. 1977. Standard aeration for gas-sterilized plastics. J Hyg Camb, 79: 225-32.

c. Most 1960s and 1970s operations had evolved to storing the sterilized materials during degassing in a separate room from chamber operations, while operations in earlier decades had chamber operations and sterilized material stored in the same workspace. In the 1950s and 1960s, sterilizer operators would be expected to have higher exposures than in the 1970s because there was one (or no) air washes and the sterilized pallets with high residue levels were often stored in the same room as the chambers.

d. Systematic application of forced and efficient ventilation where sterilizers were operating and where treated pallets were stored was rare or absent prior to the mid-1970s.

e. The period of degassing of sterilized materials was generally about 7 days during the mid-1960s and 1970s, but was ≤1 day in earlier decades. This indicates that the levels of residues in the sterilized materials and, hence, exposures were consistently high in earlier decades.

f. Although with increasing time prior to the mid-1970s sterilization operations involved smaller sterilizers (i.e., having smaller sterilizer chamber volumes), sterilizer operations involved less mechanized or non-mechanized processes, less- or non-ventilated chamber and storage operations, more leaky EO containment during sterilization, and more direct operator exposure to EO vapor (e.g., during change of filters contacting liquid EO and manual connection/disconnection of EO tanks)—factors that likely acted jointly to generate EO exposures to sterilizer operators and other related workers that were greater prior to the late 1970s than during later periods.

g. According to interviewed operators with decades of experience in the EO sterilization industry, concentrations of EO applied in sterilizers currently and since the late 1970s (400–600 mg/L) have been lower by a factor of roughly 1.5 than those applied during earlier decades, and resulting chamber concentrations of EO upon opening of sterilizer chamber doors (which at that time were not actively ventilated) thus are likely to have been equal to or (with increasing likelihood going back further in time) greater than those that occurred during 1978.

Each of these factors taken alone or in combination indicate that, compared to the sterilization worker environment starting in 1978, when technology improvements and regulatory controls were introduced with increasing frequency and stringency, it is highly probable that greater EO concentrations occurred in the sterilization worker environment from the mid-1960s to the late 1970s. Moreover, it is virtually certain that even greater EO concentrations occurred in the sterilization worker environment prior to the mid-1960s, contrary to trends in occupational exposures during those times that were extrapolated using the NIOSH statistical regression model.

The new information summarized above confirms that the SAB's concern was not effectively addressed by the IRIS Program, and therefore all assessments of EO cancer risk derived using NIOSH epidemiological study data are potentially confounded by greater magnitudes of uncertainty than are stated in the EO IRIS Assessment. These assessments are based on historical extrapolations of occupational exposures prior to the late-1970s produced by the NIOSH regression model and thus necessarily depend on the accuracy and reliability of those extrapolations. This major source of uncertainty in the EO IRIS Assessment is a key defect.

6. **Comparisons of relative reliability made between the NIOSH and UCC studies are inaccurate. These comparisons were a key basis upon which the IRIS Program rejected the UCC Study as a source of epidemiology study data for cancer risk assessment. The EO IRIS Assessment does not acknowledge and appropriately consider limitations of the NIOSH exposure assessment posed by low extrapolations of NIOSH cohort exposures to EO prior to the late 1970s without any corroborating data or any supporting engineering/process considerations derived from or directly relevant to that period of time**.

The EO IRIS Assessment argues inaccurately that the UCC exposure assessment was "too crude" to be used for exposure-response analysis (see Table 1). To the contrary, Greenberg et al. (1990) describe their categorization of departments into "high," "medium," and "low" categories based on a detailed reconstruction of processes using records and interviews of older employees.[43] The categorization was validated using frequencies of visits to the medical department for acute over exposures. The UCC exposure assessment was expanded to include individual exposure estimates, as described in detail by Swaen et al. (2009).[44] All such efforts associated with epidemiology studies require assumptions and involve uncertainties.

---

[43] Greenberg HL, Ott MG, Shore RE. 1990. Men assigned to ethylene oxide production or other ethylene oxide related chemical manufacturing: A mortality study. Br J Ind Med, 47: 221-30.

The UCC study, however, includes actual UCC data based on monitoring data from the UCC Texas plant with very similar operations from as early as 1957. Estimates for the 1940-1956 period are based on the published literature for companies using a similar process for EO production. The greatest uncertainty is for 1925-39; however, only 4.8% of the cohort worked during that period. In contrast, **approximately 70% of the NIOSH cohort had workplace exposures prior to 1978, the period of unverified exposure estimates.**

The EO IRIS Assessment's criticism of the UCC approach, i.e., it includes data from a comparable plant that was not part of the cohort, is biased because NIOSH also used exposure data from plants that were not included in the cohort. The fact that UCC-cohort exposures estimated between 1957-1973 are based on contemporary actual exposure measurements obtained from a very similar plant is a major advantage (and certainly not a deficiency) of the UCC approach relative to the NIOSH study.

In contrast, critical limitations and uncertainties associated with NIOSH's statistical regression modeling for the period prior to the late 1970s (based entirely on a fit obtained to data gathered only starting in the late 1970s, since no actual measurements of EO exposure were available for the NIOSH cohort prior to that time) are not accurately characterized or even meaningfully acknowledged in the EO IRIS Assessment or in related NIOSH publications. For example, Hornung et al. (1994) did not reveal that their approach resulted in lower, rather than higher, exposures over the entire period addressed prior to the late 1970s, with no exposures prior to 1978 exceeding those that occurred in and also were reliably estimated for 1978. As noted above, the pattern predicted by the NIOSH statistical regression model conflicts with what is known about early processes in the sterilant industry, and was characterized as "surprising" and "unrealistic" by the SAB.

The EO IRIS Assessment is highly misleading because what it refers to as NIOSH statistical regression model "validation" was done only for its post-late-1970s predictions, since no earlier EO-measurement data were available. Model extrapolations of historical EO exposure prior to the late 1970s were conjectural, relying entirely on putative explanatory power of a regression model fit to EO-measurement data that, as acknowledged by Hornung et al. (1994), exhibited a steeply declining pattern of EO exposures over time post-1977 due to regulatory concerns and EO-control measures that simply did not exist previously. New information described above confirms that the NIOSH exposure estimates for periods prior to the late 1970s are substantially and unrealistically low, and therefore are likely to have biased all assessments of EO cancer risk that relied only on NIOSH cohort study data. Moreover, the IRIS Program has failed to investigate whether such bias may render assessments of EO cancer risk unreliable.

---

[44] Swaen GM, Burns C, Teta JM, Bodner K, Keenan D, Bodnar CM. 2009. Mortality study update of ethylene oxide workers in chemical manufacturing: a 15 year update. J Occup Environ Med, 51(6): 714-23.

7. The EO IRIS Assessment relies solely on the NIOSH study of sterilant workers and fails to incorporate the important findings from the UCC study of workers in EO producing and using operations. The IRIS Program considered and characterized three factors in its selection of the NIOSH study: cohort size, exposure data, and confounding. Based on these factors, the IRIS Program dismissed the UCC study as a basis for EO cancer risk estimation. In considering cohort size, the IRIS Program ignored the most important comparison—the number of lymphohematopoietic tissue cancers, not the total cohort size.

As discussed in detail in the other sections, the NIOSH study does not have superior exposure data compared to the UCC study, so both studies have comparable applicability to risk assessment.

Cohort size is only one factor in assessing study informativeness. The most important factor is the number of events of interest, which for a mortality study is dependent on length of follow up and percent deceased. The most recent published study of the UCC cohort reports a sizeable number of deaths due to leukemia and lymphomas, comparable to the events among males in the NIOSH study that would make a meaningful contribution to the number of events for an exposure-response analysis.[45] Despite the smaller number of male workers in the UCC study, they have been followed for a longer period of time (37 yr on average compared to 25 yr for the NIOSH study) and include 51% deceased compared to 19% of the much younger NIOSH sterilant population. The EO IRIS Assessment criticizes the sample size in the UCC cohort, noting (erroneously) “only” 27 LHC cancers and 12 leukemias; the correct number of leukemias is 11 (EPA interchanged the numbers of leukemia and NHL deaths). However, the EO IRIS Assessment does not also note the male population of the NIOSH study had 37 LHC cancers and only 10 leukemias. Furthermore, no substantive criticisms of the NIOSH study appear in the EO IRIS Assessment, when in fact there are major uncertainties with respect to the NIOSH exposure estimates as described in detail above.

The EO IRIS Assessment raises concerns about confounding in the UCC study because of the presence of multiple chemicals in the workplace. This source of bias would only be expected when analyses yield positive findings, i.e., increases that may not be attributed to EO but to other chemicals. This, in fact, was identified by Greenberg et al. (1990), which reported an increase in leukemia and pancreatic cancer that was found to be attributable to exposures to one or more chemicals in the ethylene chlorohydrin production unit that was characterized as a "low" EO department. The 278 workers involved in that department were removed from the cohort and separately analyzed in a companion publication,[46] which verified increased risk observed by Greenberg et al. (1990). The remaining EO workers did not exhibit cancer increases in subsequent updates.[47] The three central reasons cited in the EO IRIS Assessment for excluding the UCC study are not defensible as explained above, and therefore indicate a biased preference for using the NIOSH study as a sole basis for EO cancer risk estimation.

---

[45] Swaen et al. (2009).

In addition, the EO IRIS Assessment diminishes the value of the most recent UCC cohort study claiming they were followed so long that background rates of lymphoid tumors would be so large as to miss increased risks due to EO. The important factor is to have sufficient time since first exposure (latency). The 37 yr. average follow-up of Swaen et al. (2003) is not excessive in light of the fact that the most recent hires (1988) have 15 yr. follow-up at most. It is desirable to have 20-25 yr. follow-up for a cancer outcome of interest and even longer when exposures are lower as they were post-1976. Furthermore, there were two earlier studies of this cohort (Greenberg et al., 1990 and Teta et al., 1993) when the cohort was younger, which failed to identify EO-related cancer increases. These studies examined the findings by hire date, duration of exposure, time since first exposure and performed comparisons to the non-exposed chemical workers adjusting for age. It is implausible and speculative that the aging of the cohort masked significant EO-related cancer increases.

The UCC study should have been incorporated in both the hazard characterization and the exposure-response analysis. Consequently, the IRIS Program's handling of these key issues—cohort size, exposure estimation, and confounding—is incomplete, inaccurate, and biased.

8. **The use of the supralinear spline model for the lymphoid and breast cancers in the final EO IRIS Assessment is based on an invalid statistical analysis. Because the analysis did not correctly calculate degrees of freedom associated with that fitted model, it contains erroneous measures of absolute and relative goodness of fit of that model. When both the p-values and Akaike**

---

[46] Benson LO, Teta MJ. 1993. Mortality due to pancreatic and lymphopoietic cancers in chlorohydrin production workers. Br J Ind Med, 50: 710-16.

[47] Teta MJ, Benson LO, Vitale JN. 1993. Mortality study of ethylene oxide workers in chemical manufacturing: A 10 year update. Br J Ind Med, 50: 704-09; Swaen et al. (2009).

Information Criterion (AIC) values characterizing fit quality are corrected, the supralinear spline model does not fit the NIOSH lymphoid tumor data statistically significantly better than the log-linear Cox model.

The EO IRIS Assessment justifies why it does not account for the degrees of freedom by citing the 2015 SAB Review: "The knot is preselected and is not considered a parameter in these analyses, consistent with the SAB's concept of parsimony (SAB, 2015)."[48] However, the concept of parsimony is a preference for a simpler model with fewer estimated parameters when fitting and evaluating a single model. The SAB did not direct EPA to violate well founded and widely accepted statistical practice by ignoring the fact that a particular parameter (in this case, the knot of a bi-linear spline model) of a spline model was actually estimated when defining the total number of its estimated parameters, when comparing the goodness of fit of that spline model to another model (such as a log-linear model) that involves no estimated knot.[49]

The EO IRIS Assessment indicates to fit particular supralinear spline models, their "knots were obtained by doing a grid search by increments of 100 ppm x days and then interpolating where appropriate."[50] In other words, the knot of the final supralinear spline model selected was indeed an additional estimated (in this case, numerically optimized) parameter, standard statistical model-fitting procedures always require that p-values be evaluated for a goodness-of-fit statistic only after subtracting one degree of freedom ***for each one of the total number of parameters (a number typically denoted as k) that are estimated when fitting a model***, regardless of how such parameters are estimated.

Failure to follow this procedure always results in an erroneously inflated "p-value" for goodness of fit (only a model with a p-value for goodness-of-fit larger than 0.05 is typically considered acceptable), and thus also in an underestimated value of a corresponding AIC used to compare goodness of fit of different models (a model with a smaller AIC value is preferred, and AIC is defined as twice the sum of *k* [defined above] and a fit-specific positive quantity). If the proper procedure is not followed to define total degrees of freedom (*k*), the result is a p-value indicating a fit that is better than actually is the case (i.e., a p-value indicating that deviations between a fitted model and the observed/modeled data are more likely to have occurred by chance alone than actually is the case), and consequently also an AIC value that misrepresents a model's goodness of fit relative to that of another model for which degrees of freedom (*k*) are defined properly.

---

[48] EO IRIS Assessment, Appendix D, at D-6.

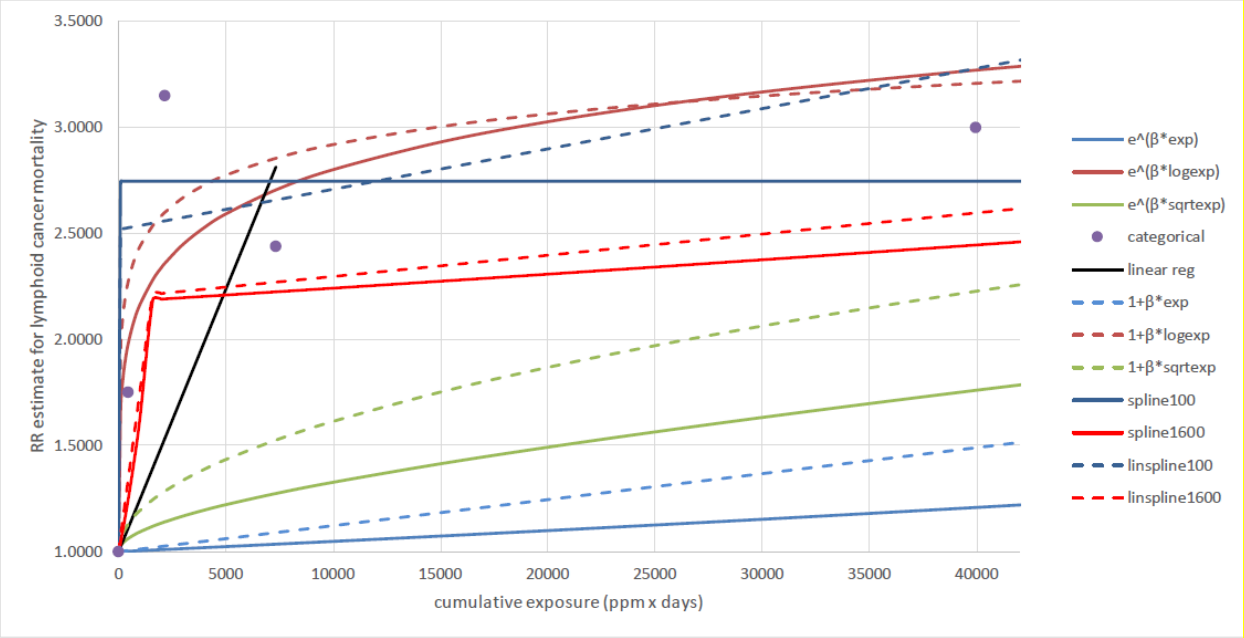
[49] The EO IRIS Assessment quotes the SAB as follows: "in some settings the principle of parsimony may suggest that the most informative analysis will rely upon fixing some parameters rather than estimating them from the data. The impact of the fixed parameter choices can be evaluated in sensitivity analyses. In the draft assessment, fixing the knot when estimating linear spline model fits from relative risk regressions is one such example." Appendix D, at D-6, note 11.

[50] EO IRIS Assessment, Appendix D, Table D-27, note a.

By ignoring this statistical procedure for its supralinear spline model fit, the EO IRIS Assessment artificially and erroneously inflates the p-value and reduces the AIC value that was used to compare that model to those of other models being compared for which degrees of freedom were defined correctly. **When both the p-values and AIC values are corrected, the selected supralinear spline model does not fit the NIOSH lymphoid tumor data statistically significantly better than the log- linear cumulative model (see Appendix 1).**

9. The selection of the supralinear spline model for the lymphoid tumors is also based on misleading illustrations of "visual fits" that do not convey either the actual data that were fit or the relative goodness of fit to these data of log-linear and supralinear spline models. Only in a footnote does the IRIS Program acknowledge that the visual comparison misrepresents the log-linear model being compared. Consequently, and erroneously, the fit to the data appears far worse than the supralinear spline model. The data plotted in that figure also were summary data that misrepresent the true magnitude of the scatter of the data that were used for model fitting.

The EO IRIS Assessment visually represents alternative models considered in relation to data used for model fitting in Figures 4-3 through 4-8, explaining that "to facilitate a visual comparison of the models, select models are replotted against the categorical data in deciles." Figure 4 below reprints Figure 4-3 from the EO IRIS Assessment and illustrates the incorrect basis for the conclusion that the NIOSH exposure-response is supralinear and that only models that are supralinear have good visual fit to the data.



e^(β*exp): RR = $e^{(\beta \times \text{exposure})}$; e^(β*logexp): RR = $e^{(\beta \times \ln(\text{exposure}))}$; e^(β*sqrtexp): RR = $e^{(\beta \times \text{sqrt}(\text{exposure}))}$; categorical: RR = $e^{(\beta \times \text{exposure})}$ with categorical exposures, plotted at the mean cumulative exposure; linear reg: weighted linear regression of categorical results, excluding highest exposure group (see text); 1 + β*exp: RR = 1 + β × exposure; 1 + β*logexp: RR = 1 + β × ln(exposure); 1 + β*sqrtexp: RR = 1 + β × sqrt(exposure); spline100(1,600): Two-piece log-linear spline model with knot at 100 (1,600) ppm × days (see text); linspline100(1,600): Two-piece linear spline model with knot at 100 (1,600) ppm × days (see text). (Note that, with the exception of the categorical results and the linear regression of the categorical results, the different models have different implicitly estimated baseline risks; thus, they are not strictly comparable to each other in terms of RR values, i.e., along the *y*-axis. They are, however, comparable in terms of general shape.)

Source: Steenland reanalyses for males and females combined; see Appendix D (except for linear regression of categorical results, which was done by EPA).

Figure 4-3. Exposure-response models for lymphoid cancer mortality vs. occupational cumulative exposure (with 15-year lag).

Figure 4. Figure 4-3 from the EO IRIS Assessment using categorical data (solid purple points) to compare the visual fits of the different models, including the selected two-piece log-linear-spline model (dashed red curve) and the standard Cox log-linear regression model (solid blue curve).

Figure 4-3 misrepresents the relative quality of true visual fits to the EO IRIS Assessment's preferred supralinear spline model compared to the more parsimonious log-linear Cox regression model in two important ways. First, Figure 4-3 plots data points that represent categorical data aggregated into quartiles (filled purple points in Figure 4, above) instead of the actual individual cases modeled. This comparison was used in earlier drafts of the IRIS Assessment when the 2014 draft EO IRIS Assessment modeled those categorical aggregated or summary data. However, when the final EO IRIS Assessment followed the SAB's recommendation to model individual cases, the data plots were not corrected accordingly to show the true magnitude of data scatter in relation to fitted models.

Second, the IRIS Program acknowledges in a footnote to Figure 4-3 that "the various models have different implicitly estimated baseline risks; thus, they are not strictly comparable to each other in terms of RR values (i.e. along the y-axis). They are, however, comparable in terms of general shape." It is not transparent, however, that these graphs cannot be used at all to compare some of the models shown in a valid way. In particular, the lower log-linear model fit shown (the solid blue "line" that appears to go through the origin of the plot shown in Figure 4-3) appears to provide a very poor fit to the cloud of individual data through which that model passes, because the place where that model is shown to intersect the y-axis was artificially forced (in that figure) to intersect the value of 1 along the y-axis, when in fact that model does actually pass centrally through the cloud of actual raw data to which it was fit. That is, although both the EO IRIS Assessment's preferred model and the log-linear model do more or less centrally pass through the cloud of data to which these models were fit, Figure 4-3 misleads the reader by showing a relatively poor fit of the simpler (i.e., more parsimonious) log-linear model compared to the more complex supralinear spline model that was selected in the EO IRIS Assessment.

Figure 5[51] more accurately compares the supralinear spline model (red dashed curve) and the standard Cox log-linear regression model (solid blue curve). The latter model is the approach used by Valdez-Flores et al. (2010) to fit the NIOSH, UCC, and combined NIOSH+UCC study data for lymphoid tumors. In Figure 5, the baseline (zero-exposure) value of hazard rate (HR) to which the log-linear model was fit is set equal to the same baseline HR as that estimated using the supralinear spline model. Therefore, Figure 5 shows more accurately than Figure 4 that the supralinear spline model fits the data no better than standard Cox log-linear regression model.

---

[51] Figure 5 improves comparison along the y-axis by dividing model-estimated values of hazard rate (HR) ratio by the baseline HR of the individual categorical cases (thus making an apples-to-apples comparison), and uses a logarithmic scale to improve comparison of the linear difference between the fitted models and observed values of relative risk measured as hazard rate ratio (RR). In Figure 4, RR values greater than one appear disproportionally more distant from 1 than RR values less than one, because of the linear RR scale used in that figure. RR values greater than one can be as large as infinity, but RR values less than one cannot be less than 0. In contrast, values of Ln(RR)—i.e., values of RR plotted on a logarithmic scale—as shown in Figure 5 can be as large as infinity and as small as minus infinity (see Appendix 1).

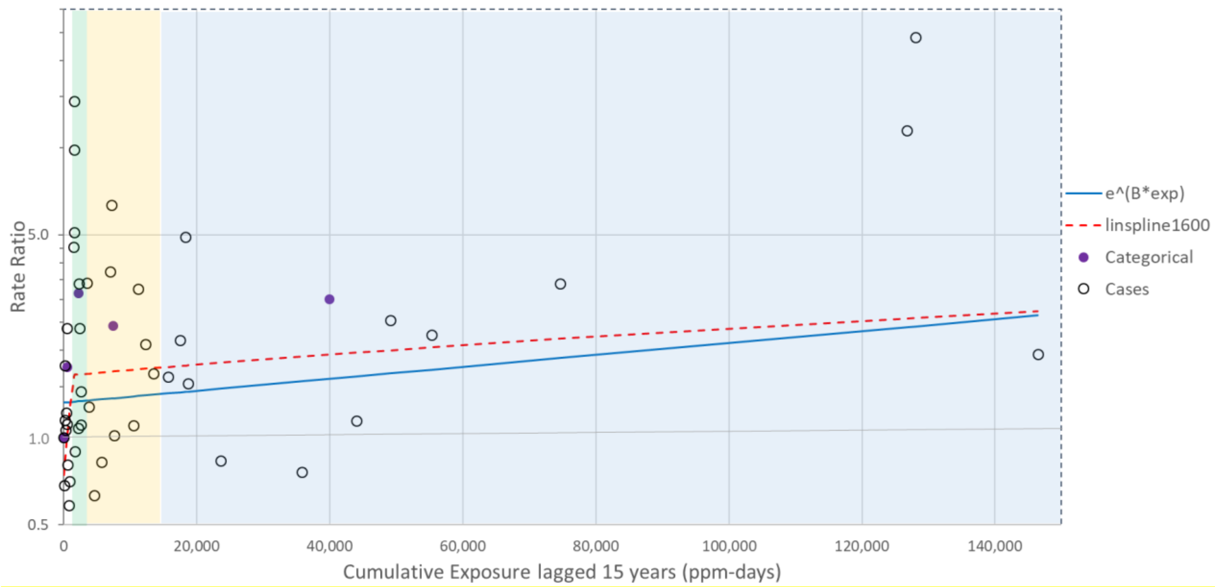


Figure 5. Apples-to-apples comparison of the EO IRIS Assessment's preferred supralinear spline model (red dashed curve) and the log-linear Cox proportional hazards model (solid blue curve), plotted in relation to categorical data (solid purple points) from Figure 4 together with corresponding actual (raw/individual-level) data to which these models were fit (open points).

The misleading plots of categorical data in the EO IRIS Assessment were a key justification for its rejection of the standard Cox log-linear proportional hazards model in favor of a supralinear exposure-response relationship, as indicated in Table 4-14 of the EO IRIS Assessment.

10. The selection of a spline model as the preferred model for EO cancer risk estimation assumes a supralinear increase in tumor response in the low-dose exposure region with a subsequent plateauing of response at higher exposures. The body of cancer epidemiologic studies, including the NIOSH studies, does not support such a pattern of risk. While certain NIOSH sub-analyses suggest increases in male lymphoid tumors and female breast cancers, the findings are limited to the highest cumulative exposure groups, not the lowest.

Steenland et al. (2003) state, "Exposure-response data do suggest an increased risk … for those with higher cumulative exposures to ETO."[52] The authors also say, "The dip in the spline curve in the region of higher exposures suggested an inconsistent or non-monotonic risk with increasing exposure." The default expectation for a genotoxic carcinogen would be this pattern of monotonically increasing risk in relation to exposure, which is why the authors call it "inconsistent." The EO IRIS Assessment notes that it is not unexpected to have fluctuations in exposure-response curves due to random variation, yet in the exposure-response section the IRIS Program models such plausibly random fluctuation using a supralinear response model.

---

[52] Steenland K, Whelan E, Deddens J, Stayner L, Ward E. 2003. Ethylene oxide and breast cancer incidence in a cohort study of 7576 women (United States). Cancer Causes Control, 14: 531-39.

The EO IRIS Assessment cites Mikoczy et al. (2011)[53] to support the use of the supralinear spline model for breast cancer: "Although the reason for the observed supralinear exposure-response relationship is unknown, it is worth noting that the results of the Swedish sterilizer worker study reported by Mikoczy et al. 2011, …support the general supralinear exposure-response relationship observed in the NIOSH study."[54] However, Mikoczy et al. (2011) studied a low-exposure population that exhibited a significant increase in breast cancer incidence only when analyzed using an internal analysis comparing more-highly exposed to low-exposed workers, and exhibited no such significant increase in a corresponding external analysis involving comparison to matching members of a general population. The explanation for this anomaly lies in the dramatic and (as indicated by Mikoczy et al., 2011) statistically significant deficit of breast cancers in the low exposure group of the internal comparison; because in the internal comparison that low-exposed group was used as the referent group, the two higher exposure groups being compared showed significantly higher rates breast cancer relative to that lower-exposed group.

It might be argued that the non-representative and significantly low rate of breast cancer incidence exhibited by the low-exposure group used for internal comparison simply reflects a Healthy Worker Effect (HWE). However, the breast cancer rate for that group was remarkably low (only about half that of the reference population group of age-matched Swedish women used), and there is no HWE specific to breast (or to any other type of) cancer in Swedish female workers.[55] Thus, the EO IRIS Assessment does not accurately acknowledge and address the problematic nature of the internal-comparison reference group that served as the basis for results of internal comparisons of breast cancer incidence reported by Mikoczy et al. (2011).

---

[53] Mikoczy Z, Tinnerberg H, Jonas Björk J, Albin M. 2011. Cancer incidence and mortality in Swedish sterilant workers exposed to ethylene oxide: updated cohort study findings 1972–2006. Int J Environ Res Public Health, 8: 2009-19.

[54] EO IRIS Assessment, at 4-71.

[55] Gridley G, Nyren O, Dosemeci M, Moradi T, Adami HO, Carroll L, Zahm SH. 1999. Is there a healthy worker effect for cancer incidence among women in Sweden? Am J Ind Med, 36(1): 193-99.

The EO IRIS Assessment's extra risk estimate suggests a highly potent carcinogen. This is contrary to epidemiology findings which show overall weak positive findings (see Appendix 2). While interest has centered on leukemia, other blood related malignancies, and recently on breast cancer, there are numerous inconsistencies among the studies; elevated risks above background, in isolated studies, are of small magnitude; and there is an absence of a clear exposure-response for any specific cancer type. The most informative studies are the NIOSH (Steenland et al. 2003, 2004) and UCC studies (Swaen et al. 2009), which are studies of comparable utility for risk assessment purposes. These epidemiology studies do not support supralinearity (high risk at low exposures). Certain NIOSH subanalyses showed increase for males only (lymphoid tumors) in the highest (not the lowest) cumulative exposure groups. Extended follow up of chemical workers, UCC and others, and sterilant workers show little, if any, increases. The epidemiological evidence does not support the RSC of 0.1 ppt, which suggests a highly potent carcinogen.

11. The use of a supralinear spline model for cancer risk estimation is inconsistent with the assumed mode-of-action of EO toxicity and tumorigenicity. Such a model predicts higher risk at low exposures compared to risks predicted at higher exposures, which is contradicted by the well-understood mode of action of EO in experimental animals and humans as described in the EO IRIS Assessment. Thus, the EO IRIS Assessment relies on human cancer risk estimates based on spline-model dose-response extrapolations that are internally inconsistent with its own evaluation of the mode of action of EO. The mean air concentration equivalent to the endogenous concentration in non-smoking humans with no known EO exposures is 1.9 ppb (range 0.13-6.9 ppb; continuous), which is 19,000 times greater than the EO IRIS RSC of 0.1 ppt. An alternative LEC (1/million) of 0.5-1.2 ppb is a more pragmatic, science-based approach for EO risk assessment.

As a direct acting DNA- and protein-reactive toxicant, the high-level toxicological and cancer mode of action of EO importantly predicts a *sublinear* increase in dose-response at low exposures and an associated dose-disproportionate *increase* in toxicity at higher EO doses.[56] This expected dose-response pattern is due to attenuation of low-dose EO toxicity mediated by intervention of key detoxification pathways (EO conjugation with glutathione and enzymatic hydrolysis to oxidized metabolites; repair of EO-induced DNA adducts), and an associated dose-disproportionate (supralinear) increase in toxicity at higher doses due to saturation of those same pathway(s) as the EO dose increases, as summarized below in Figure 6.

---

[56] Kirman and Hays (2017).

The EO IRIS Assessment describes and supports this projected EO mode of action and its implications for the shape of the cancer dose response in the low- to high-dose regions as follows:

> [E]PA considers it *highly plausible that the dose-response relationship over the endogenous range is sublinear* (e.g., that the baseline levels of DNA repair enzymes and other protective systems evolved to deal with endogenous DNA damage would work more effectively for lower levels of endogenous adducts), that is, that the *slope of the dose-response relationship for risk per adduct would increase* as the level of endogenous adducts increases.[57]

The EO IRIS Assessment's analysis of the EO mode of action emphasizes that the dose-response is highly likely ("highly plausible") to be *sublinear* "over the endogenous range" of internal EO doses that result from well-characterized endogenous production of EO secondary to metabolism of ethylene originating from normal biological processes.

Exploiting the well-defined linear relationship between exogenous EO exposure and systemic hemoglobin adducts in humans, Kirman and Hays (2017) estimate that the contribution of endogenously generated EO exposures to the overall systemic dose of EO is substantially greater than the 0.1 ppt exogenous EO exposure projected by the EO IRIS Assessment as resulting in a 1 x 10^{-6} cancer risk in humans. A meta-analysis of 661 non-smoking individuals not exposed to external EO indicated that endogenous background EO exposures are equivalent to a mean external exogenous EO exposure of 1.9 ppb (range 0.13-6.9 ppb). This "endogenous equivalent" contribution to the overall systemic EO dose is 19,000 times greater than the 0.1 ppt exogenous EO one-in-a-million risk dose estimated by the EO IRIS Assessment.

It is clear that even a 1000-fold increase in exogenous EO exposures above 0.1 ppt would only approach the low end of the total systemic EO dose contributed by endogenous EO generation. Any contributions of exogenous EO to cancer risk below this low-end endogenous dose would not be detectable within the likely day-by-day intra- and inter-individual variability (0.13-6.9 ppb) associated with normal endogenous EO exposure loads.

---

[57] EO IRIS Assessment, at 4-95 (emphasis added).

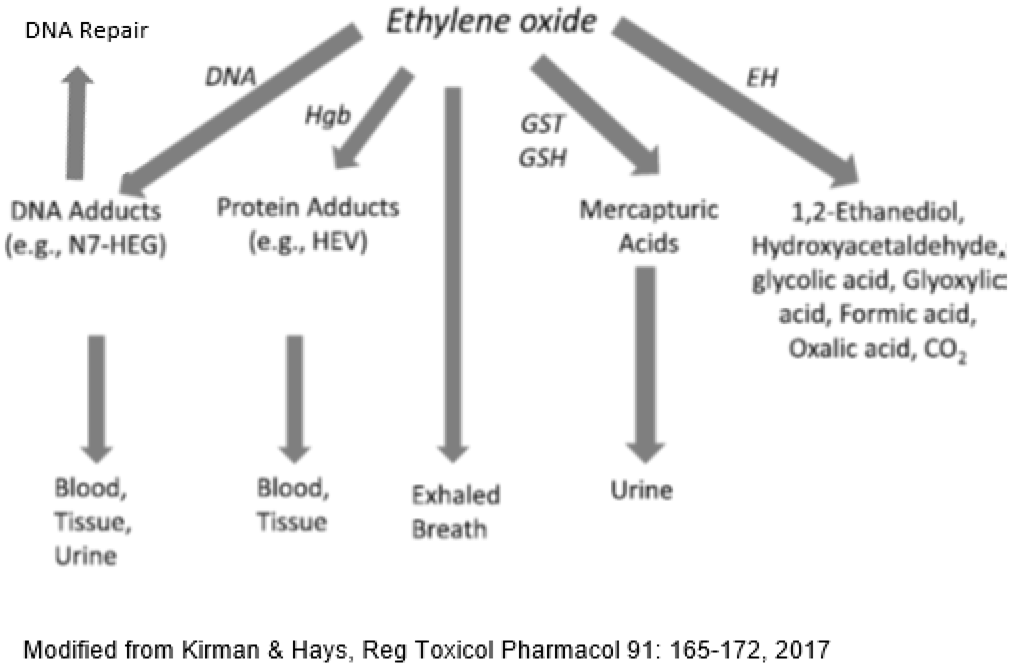

Ethylene oxide
DNA Repair
DNA
Hgb
GST
GSH
EH
DNA Adducts (e.g., N7-HEG)
Protein Adducts (e.g., HEV)
Mercapturic Acids
1,2-Ethanediol, Hydroxyacetaldehyde, glycolic acid, Glyoxylic acid, Formic acid, Oxalic acid, CO2
Blood, Tissue, Urine
Blood, Tissue
Exhaled Breath
Urine
Modified from Kirman & Hays, Reg Toxicol Pharmacol 91: 165-172, 2017

Figure 6. EO metabolism (adapted from Kirman and Hays, 2017).

Kirman and Hays (2017) also recognize that increased EO hemoglobin adducts associated with smoking provided an opportunity to further check the EO IRIS Assessment's supralinear model predictions that moderately low external EO exposures realistically contribute to increased cancer risks. A meta-analysis of 379 smokers not otherwise exposed to EO found that smoking increased EO exposures approximately 10-fold above the endogenous equivalent dose for background (non-EO exposed) individuals (mean background endogenous equivalent exposure = 1.9 ppb; mean smoker exposure = 18.8 ppb). The spline-model relied on by the EO IRIS Assessment predicts that the moderate increase in EO exposure associated with smoking would result in a detectable increase in lymphohematopoietic and breast cancers. However, this expectation is not met despite the very large smoking cohort.

Kirman and Hays (2017) note that smoking has been causally associated only with one subtype of lymphohematopoietic cancer, acute myeloid leukemia (AML). Not only is this cancer not increased in the NIOSH occupational cohort specifically exposed to higher doses of EO than those resulting from smoking, but Valdez-Flores et al. (2010), using a non-spline-based risk model, also demonstrate a statistically significant negative slope between cumulative exposure to EO and AML in that same NIOSH cohort. Kirman and Hays (2017) also observe that evidence of a causal relationship between smoking and breast cancer is considered only as suggestive and not sufficient. Thus, projections of low-dose elevations in specific EO-associated cancer risks based on spline model extrapolations from relatively high occupationally-exposed individuals are not consistent with cancer outcomes in the much larger smoking cohort experiencing moderately elevated EO exposures.

Kirman and Hays (2017) also address the concern that any additional exogenous EO exposures above background, regardless of how small, represent a plausible contribution to increased cancer risks. They conclude that the approximate four order of magnitude disparity between EO endogenous exposures (mean = 1.9 ppb) and EPA projected increased risk at exposures greater than 0.1 ppt "creates a signal-to-noise issue [in the biological plausibility of tumor outcomes] when exogenous exposures fall well below those consistent with endogenous exposures. In such cases, small exogenous exposures may not contribute to total exposure or to potential effects in a biologically meaningful way."

Recently, Calabrese (2018)[58] offers additional insight into the lack of plausibility of additivity to background of risks associated with low (and particularly less than background) exposures to EO. Calabrese reports that the mutational spectra of K-ras in EO-induced lung and Harderian gland tumors, and H-ras and p53 in mouse mammary tumors, were not at all similar to mutational spectra of these same tumors in control mice from the EO studies. These molecular-level data indicate that the mode of action of generation of control (background) tumors differs substantively from those originating from exogenous EO-exposed animals, even though control animals experience significant endogenous EO exposures. Thus, these data stand in contrast to the assumption of additivity to background that presumes that chemically-induced elevation of background tumors that are otherwise pathologically similar to chemically-induced tumors must share common mode(s) of action reviewed by Calabrese (2018).

The potential for additivity to background also is not supported by a comparison of total endogenous EO-specific DNA adducts in spleen, liver and stomach of rats relative to adducts in these same tissues resulting from a thousand-fold range of EO intraperitoneal doses (0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05 and 0.1 mg/kg/day; 0.1 mg/kg/day approximately equivalent to a 1 ppm 6 hr/day EO inhalation exposure).[59] Importantly, Marsden et al. (2009) also emphasize that the increase in adducts associated with exogenous EO were not statistically significant at any dose with the exception of adducts in liver in rats administered 0.05 mg/kg/day, suggesting that exogenous adducts may not present any additional risk over endogenous adducts over this range of EO doses (i.e., additivity to background). Interestingly, endogenous DNA adducts were statistically increased in spleen and liver at the 0.05 and 0.1 mg/kg/day EO, indicating that higher EO doses alter internal biological processes leading to increased potential for endogenous EO formation.

---

[58] Calabrese EJ. 2018. The additive to background assumption in cancer risk assessment: A reappraisal. Envir Res, 16: 175-204.

[59] Marsden DA, Jones DJ, Britton RG, Ognibene T, Ubick E, Johnson GE, Farmer PB, Brown K. 2009. Dose-response relationships for N7-(2-hydroxyethyl)guanine induced by low-dose [^{14}C]ethylene oxide: evidence for a novel mechanism of endogenous adduct formation. Cancer Res, 69(7): 3052-59.

Further investigations demonstrated that the high-dose-specific in endogenous-only adducts may have been secondary to increased oxidative stress. Both the high level of background endogenous adducts and high-dose specific increases in endogenous-only EO adducts further supports the authors' conclusion that "if the compound [EO] is produced endogenously, low doses of exogenous exposure may be overwhelmed by the background levels, leading to no detectable statistically significant increase in risk due to the external exposure." This conclusion (see Figure 6) is entirely consistent with the analyses developed by Kirman and Hays (2017) in which endogenous EO equivalent exposures in humans (mean = 1.9 ppb) are estimated as being 19,000 times higher than the exogenous EO dose of 0.1 ppt presenting a one-in-a-million cancer risk from spline-model low-dose extrapolation.

An alternative LEC (1/million) of 0.5-1.2 ppb is within the range of endogenous EO levels. Taking into account the biological mode of action and the endogenous EO equivalent exposures in humans, this approach is more plausible and science-based than the EO IRIS assessment.

12. The statistical, epidemiological and biological evidence does not support the selection of supralinear spline models to fit the NIOSH study data in the EO IRIS Assessment. A more scientifically sound conservative alternative is to use the Valdez-Flores et al. (2010) approach, which incorporates all the available data from the two strongest human studies (NIOSH and UCC). This approach has been adopted by the Scientific Committee on Occupational Exposure Limits.

As described in previous sections, the selection of the supralinear spline model is based on incorrect statistical analysis and biased evaluation of the NIOSH exposure modeling relative to the UCC exposure estimates. Furthermore, the epidemiological evidence and biological mode of action do not support the supralinear spline model. A more scientifically supportable approach is that published by Valdez-Flores et al. (2010), who make full use of the available data from both the NIOSH and UCC cohorts. The effect was modeled as a standard Cox proportional log-linear hazards model (i.e., exponentiated linear) function of cumulative EO exposure (ppm-days) treated as a continuous variable.

The EO IRIS Assessment focuses the cancer risk assessment on lymphoid tumors (defined by NIOSH as including non-Hodgkin's lymphoma, lymphocytic leukemia and multiple myeloma) and breast cancer incidence. The weight of evidence does not support breast cancer as an endpoint for risk assessment (see Appendix 2). Therefore, our analysis focuses on the mortality data for lymphohematopoietic (LH) tissue cancers including leukemia (and specific myeloid and lymphocytic leukemia), non Hodgkin's lymphoma (NHL), multiple myeloma (MM) and "lymphoid" cancers (a grouping developed in Steenland et al. (2004) that included NHL, MM, and lymphocytic leukemia).

Valdez-Flores et al. (2010) propose a range of 1-3 ppb based on the Maximum Likelihood Estimate (MLE) of the Effective Concentrations (ECs) associated with an extra risk of one-in-a-million [EC(1/million)] (see Table 2).[60] The authors select the MLE as the most reliable data for point of departure because the Lowest Effective Concentrations LECs), the 95% lower bound on the ECs, are insensitive to the magnitude of the best estimated slope, which can be negative, yet have a positive 95% upper confidence limit resulting in a finite LEC as occurred for multiple myeloma.

Table 2: Maximum Likelihood Estimate (MLE) of the EC (1/million) and Lowest Effective Concentration (LEC)

EO type of cancer (mortality)	MLE UCC & NIOSH (ppb)	LEC UCC & NIOSH (ppb)	LEC NIOSH only (ppb)
Lymphoid	1.5	0.5	0.2
Non-Hodgkin's lymphoma	2.3	0.9	0.8
Multiple Myeloma	Negative slope, value not calculated	1.2	0.8
Leukemia	9.2	0.9	0.9
Lymphocytic Leukemia	2.4	0.9	0.9
Breast cancer	0.7	0.1	0.1

---

[60] NIOSH only provided ACC with the breast cancer mortality and not the incidence data, despite multiple requests for the incidence data. The results from the breast cancer mortality are included in Table 2 for completeness.

EO type of cancer (mortality)	MLE UCC & NIOSH (ppb)	LEC UCC & NIOSH (ppb)	LEC NIOSH only (ppb)
Range for LHC	1.5-9.2	0.5-1.2	0.4-0.9
Range for LHC and breast cancer	0.7-9.2	0.1-1.2	0.1-0.9

The MLE and LEC values reported in Table 2 are conservative values because (a) extra risk was calculated despite no statistically significant slope in the exposure-response analyses; (b) the NIOSH data was included without adjustment for likelihood of underestimation of exposures; and (c) the limited evidence of cancer risk based on the entire body of epidemiologic evidence (summarized in Appendix 2).

The EO IRIS Assessment and Valdez-Flores et al. (2010) identify several differences between the two approaches in deriving their recommended 1/million exposure levels to use as points of departure (see Table 3).[61]

Table 3: Approximate sources of differences between Valdez-Flores et al. (2010) and EO IRIS Assessment approaches

Valdez-Flores et al (2010) compared to EO IRIS Assessment	Reference	Factor
Extra risk at age 70 instead of 85 years	Valdez-Flores et al. (2010), p. 319	2.3
Different approaches to implementing age-adjusted adjustment factor (ADAF)	Valdez-Flores et al. (2010), p. 319 used an approach that adjusted the slope; EPA's cancer risk assessment guidelines (2005) use 1.66	1.66
Use of incidence background rates compared to mortality background rates in lymphoid tumor unit risk estimation (incidence/mortality ratio, $R_{i/m}$).	$R_{i/m}$ = 5.26/1.99 The EO IRIS Assessment unit risk using background lymphoid cancer incidence rates with model for lymphoid mortality data = 5.26/ppm, and unit risk using background	2.64

---

[61] *See* EO IRIS Assessment, Appendix A, at A-33 – A-35.

	mortality rates with model for lymphoid mortality data is 1.99/ppm; see Table 4-7, page 4-23; whereas Valdez-Flores et al. (2010) unit risk using background lymphoid mortality rates with model for lymphoid mortality data	

Valdez-Flores et al (2010) used well-accepted statistical principles to guide decisions about whether to include a lag period, how to calculate the degrees of freedom, and whether the MLE for the EC (1/million) can be interpolated within the lower region of the experimental data set. For example, because there was no significance between the models with and without a lag period and no clear biological plausibility for selection of a specific lag period, the more parsimonious model (no lag) was selected. In contrast, the IRIS Program tested different lag periods and knots but did not fully account for the higher degrees of freedom typically considered when different ranges of values are tested.

Valdez-Flores et al. (2010) also modeled down to 10^{-6} risk, whereas the IRIS Program modeled to 10^{-2} risk and used the LEC01 as a point of departure (POD) for linear low-dose extrapolation. Valdez-Flores et al. (2010) suggest that PODs should be within the range of observed exposures, and chose a 10^{-6} risk level because the corresponding exposure level was in the range of the observed occupational exposures (converted to equivalent environmental exposures). Thus, Valdez-Flores et al. (2010) fully used the experimental data to derive a 10^{-6} risk level.

An additional difference that is not captured in Table 3 is the EO IRIS Assessment estimates risk for both lymphoid and breast cancer, whereas Valdez-Flores et al. (2010) estimates risk for lymphoid tumors alone. As discussed above and in greater detail in Appendix 2, breast cancer is not a target of EO. The EO IRIS Assessment recognizes that magnitudes of increased risks for breast cancer were not large and implies that the evidence is weaker than that for lymphoid tumors. Despite these issues, the EO IRIS Assessment introduces breast cancer as a target organ and inappropriately develops a risk value. Uncertainties described by Steenland et al. (2003) related to the breast cancer incidence study are dismissed as unimportant. It is notable that the ratio between risk for lymphoid plus breast cancer incidence (6.06 per ppm)[62] divided by the risk for lymphoid tumor incidence alone (5.26 per ppm)[63] is only 1.15.

---

[62] EO IRIS Assessment, at 4-58.

[63] *Id.* at 4-31.

As discussed above, the NIOSH exposure assessment was not validated prior to the late 1970s and likely underestimated exposures. In contrast, the UCC exposure estimation from the 1940s to 1970s was based on actual data from similar operations during the same time period.[64] The greatest uncertainty is between 1925-1939, but only 4.8% of the UCC cohort had work history before 1940.[65] These uncertainties are no greater than the NIOSH study uncertainties and do not justify study rejection for exposure-response analysis. Both studies are well-conducted epidemiology studies with comparable power in terms of number of events for males and of comparable utility in terms of individual exposure estimates. In fact, the UCC study was originally a NIOSH study, in that it was nested within a NIOSH/UCC collaborative study of 29,000 UCC workers in the Kanawha Valley of West Virginia.[66]

The EO IRIS Assessment also criticizes Valdez-Flores et al. (2010) for not using any log cumulative exposure models which were found to be statistically significant in analyses by Steenland et al. (2004), consistent with the apparent supralinearity of the NIOSH exposure-response data. Yet, the EO IRIS Assessment also considers the log cumulative exposure model to be "problematic because this model, which is intended to fit the full range of occupational exposures in the study, is inherently supralinear …, with the slope approaching infinity as exposures decrease towards zero, and results can be unstable for low exposures."[67]

Similarly, the IRIS Program rejected other statistically significant models due to unstable results for low exposures. As noted above, the assumption of supralinearity is based on a flawed statistical analysis of its preferred-model fit and on a misleading visual comparison of invalidly overlaid models plotted in relation to categorical data grouped in quartiles instead of considering the pattern of RR for individual cases, which more realistically reveals a very noisy data cloud through which the simpler and traditionally accepted Cox proportional model fits as well as the supralinear spline model.

Crump (2005) noted that:

> Because of these potential distortions of the exposure-response shape, one should be cautious in drawing conclusions about the shape of the exposure response from epidemiological data. Since even random, unbiased errors in exposure measurement will convert a linear exposure response, and can convert sub-linear

---

[64] Swaen et al. (2009).

[65] *Id.*

[66] Rinsky RA, Ott G, Ward E, Greenberg H, Halperin W, Leet T. 1988. Study of mortality among chemical workers in the Kanawha Valley of West Virginia. Am J Ind Med, 13: 429-38.

[67] EO IRIS Assessment, at 4-10.

> response, into a seemingly supralinear shape, one should be particular[ly] cautious about concluding an exposure-response is truly supralinear. In particular, it could be inadvisable to extrapolate an observed supralinear exposure response to low exposures to predict human risk.[68]

Crump's caution is especially relevant to the NIOSH data in light of the high potential for exposure misclassification in the earlier years of the NIOSH study when there was no data to validate the NIOSH exposure model, as described above. EPA's cancer risk assessment guidelines echo this caution: "a steep slope [i.e., supralinear] also indicates that errors in an exposure assessment can lead to large errors in estimating risk."[69]

D. Conclusion

The 2014 NATA fails to meet the requirements of the IQA and the OMB and EPA Guidelines because its use of the EO IRIS Assessment is not the best available science. Therefore, the 2014 NATA risk estimates for EO should be withdrawn and corrected to reflect scientifically-supportable risk values and EPA should not use the EO IRIS Assessment's inhalation RSC of 0.1 ppt to calculate EO risk in its ongoing CAA Section 112 RTR rulemakings and other regulatory actions. As discussed above, a more reasonable and scientifically supportable approach to an exposure response analysis yields ranges for the MLE (1.5-9.2 ppb) and LEC (0.5-1.2 ppb) that are more than three orders of magnitude greater than the EO IRIS Assessment's environmental concentration associated with one-in-a-million risk.

Sincerely,

William Gulledge

William P. Gulledge
Senior Director
Chemical Products & Technology Division

Enclosures:

Appendix 1 – Statistical Issues with EPA's Calculation of p-values and AIC's for Spline Models and Linear Models in the EO IRIS 2016
Appendix 2 – Brief Summary of Epidemiological Data for EO

---

[68] Crump KS. 2005. The effect of random error in exposure measurement upon the shape of the exposure response. Dose-Response, 3: 456-64.

[69] EPA, Guidelines for Carcinogen Risk Assessment, at 3-19.

Appendix 1

Statistical Issues with EPA's Calculation of p-values and AIC's for Spline Models and Linear Models in the EO IRIS 2016

Ciriaco Valdez-Flores, Ph.D., P.E.
Professor of Practice
4073 Emerging Technologies Building
3131 TAMU
College Station, TX 77843-3131
Tel. (979) 458-2366 Fax: (979) 458-4299
e-mail: ciriacov@tamu.edu

August 23, 2018

Introduction

The document "Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide (CASRN 75-21-8) In Support of Summary Information on the Integrated Risk Information System (IRIS), December 2016" (EO IRIS 2016) has several statistical inaccuracies that play an important role in model selection and, ultimately, in the risk assessment of EtO. The exposure-response modeling of lymphoid mortality for the NIOSH study is reviewed here, and statistical pitfalls are highlighted. EPA's statistical numbers are corrected herein and new results are derived. These corrected results question conclusions drawn by EPA about model selection. Although EPA's conclusions for the other endpoints are not analyzed herein, similar statistical pitfalls must have been incurred, as the statistical pitfalls are related to the methodology that was used for all endpoints analyzed by EPA.

Table 1 reproduces Table 4-6 of EO IRIS 2016. In this table EPA to summarizes how the linear spline model with knot at 1600 ppm × days was selected to describe the relationship between lymphoid mortality rate ratio and cumulative exposures to EO. The summary in the table indicates that the model was selected because: a) adequate statistical fit; b) adequate visual fit; c) including local fit (visual) to low-exposure range; linear fit; and d) AIC within two units of lowest AIC models considered.

It can also be shown (using the likelihood ratio test -- analyses not presented here) that EPA's selected linear spline model does not fit the NIOSH lymphoid mortality data statistically significantly better (at the 5% significance level) than the nested linear model. Similarly, log-linear spline model with knot at 1600 ppm-days does not fit the NIOSH lymphoid mortality data statistically significantly better (at the 5% significance level) than the nested log-linear model. Thus, according to the following SAB recommendation on page 12, the log-linear and the linear models should be preferred over the log-linear spline and linear spline models, respectively:

> Third, the principle of parsimony (the desire to explain phenomena using fewer parameters) should be considered. Attention to this principle becomes even more important as the information in the analysis dataset becomes even more limited.

Thus, models with very few estimated parameters should be favored in cases where there are only a few events in the dataset.

Table 1. The following table has been extracted from EO IRIS 2016 Table 4-6

Table 4–6. **Models considered for modeling the exposure-response data for lymphoid cancer mortality in both sexes in the National Institute for Occupational Safety and Health cohort for the derivation of unit risk estimates**

Model[a]	*p-value*[b]	AIC[c]	Comments
Two-piece spline models			
Linear spline model with knot at 1,600 ppm × days	0.07	462.1	**SELECTED.** Adequate statistical and visual fit, including local fit to low-exposure range; linear model; AIC within two units of lowest AIC of models considered.
Linear spline model with knot at 100 ppm × days	0.046	461.4	Good overall statistical fit and lowest AIC of two-piece spline models, but poor local fit to the low-exposure region, with no cases below the knot.
Log-linear spline model with knot at 1,600 ppm × days	0.07	462.6	Linear model preferred to log-linear (see text above).
Log-linear spline model with knot at 100 ppm × days	0.047	461.8	Good overall statistical fit and tied for lowest AIC[c] of two-piece spline models, but poor local fit to the low-exposure region, with no cases below the knot.
Linear (ERR) models ($RR = 1 + \beta \times exposure$)			
Linear model	0.13	463.2	Not statistically significant overall fit and poor visual fit.
Linear model with log cumulative exposure	0.02	460.2	Good overall statistical fit, but poor local fit to the low-exposure region.
Linear model with square-root transformation of cumulative exposure	0.053	461.8	Borderline statistical fit, but poor local fit to the low-exposure region.
Log-linear (Cox regression) models ($RR = e^{\beta \times exposure}$)			
Log-linear model (standard Cox regression model)	0.22	464.4	Not statistically significant overall fit and poor visual fit.
Log-linear model with log cumulative exposure	0.02	460.4	Good overall statistical fit; lowest AIC[c] of models considered; low-exposure slope becomes increasingly steep as exposures decrease, and large unit risk estimates can result; preference given to the two-piece spline models because they have a better ability to provide a good local fit to the low-exposure range.
Log-linear model with square-root transformation of cumulative exposure	0.08	462.8	Not statistically significant overall fit and poor visual fit.

[a] All with cumulative exposure as the exposure variable, except where noted, and with a 15-yr lag.

[b] *p*-values from likelihood ratio test, except for linear regression of categorical results, where Wald *p*-values are reported. $p < 0.05$ considered “good” statistical fit; $0.05 < p < 0.10$ considered “adequate” statistical fit if significant exposure-response relationships have already been established with similar models.

[C] AICs for linear models are directly comparable and AICs for log-linear models are directly comparable. However, for the lymphoid cancer data, SAS proc NLP consistently yielded −2LLs and AICs about 0.4 units lower than proc PHREG for the same models, including the null model, presumably for computational processing reasons, and proc NLP was used for the linear RR models. Thus, AICs for linear models are equivalent to AICs about 0.4 units higher for log-linear models. No AIC was calculated for the linear regression of categorical results.

EPA's Misinterpretation of SAB Comments about the Knot of Spline Models

EPA justifies the p-values and AIC values for the linear spline and log-linear spline models in their Table 4-6 misquoting SAB's comments. In section D.3.2 of the appendices (reference), EPA states (emphasis added) "Table D-27 also presents the AIC values for the same models to facilitate comparison with the two-piece spline models, which include an extra parameter. [The knot is preselected and is not considered a parameter in these analyses, consistent with the SAB's concept of parsimony (SAB, 2015)].[14]" Their footnote 14 in the same sections states "[14] in some settings the principle of parsimony may suggest that the most informative analysis will rely upon fixing some parameters rather than estimating them from the data. The impact of the fixed parameter choices can be evaluated in sensitivity analyses. In the draft assessment, fixing the knot when estimating linear spline model fits from relative risk regressions is one such example" [page 12 of SAB (2015)]."

Although the SAB quote is accurate, the quote just a fragment of a response and is taken out of context. The full question and SAB response are as follows (emphasis added):

> *2b: For the (low-exposure) unit risk estimates, EPA presents an estimate from the preferred model as well as a range of estimates from models considered "reasonable" for that purpose (Sections 4.1.2.3 and 4.5 and Chapter 1). Please comment on whether the rationale provided for defining the "reasonable models" is clearly and transparently described and scientifically appropriate.*
>
> The SAB understands that the EPA considered four "reasonable" models for providing unit risk estimates; these all have unit risk estimates reported in Table 4-13. A few additional models are described in Tables 4-12 and 4-13, some of which could also be considered reasonable. The presentation of "reasonable" models considers model fit and some a priori (but not clearly articulated) notion about the acceptable shape of the dose-response function in the low-dose region. Because the data do not appear to conform to the a priori notion, the draft assessment also considers models based on an untransformed continuous exposure term or a linear regression of the categorical results as reasonable. However, these models do a poorer job reflecting the patterns in the data. Although much of the approach is scientifically appropriate, the SAB does not agree with all of the judgments. In order to strengthen the assessment and presentation, some modifications are suggested to the approach for comparing models and choosing which models are reasonable. The SAB recommends that

the discussion be revised to provide more clarity and transparency as well as making the disposition easier to follow. In general, discussion of statistical significance should occur in a more nuanced fashion so that important perspective about the results is not lost in the tendency to turn the statistical evidence into a binary categorization of significant vs. not significant. (This can mislead readers into interpreting a pair of results as inconsistent when their p-values, effect estimates, and 95% confidence intervals are very similar, but the two p-values happen to be on opposite sides of 0.05.) Consideration of reasonable models should address the quality of 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. Discarding a model because the fitted curve is “too steep” needs scientific justification. Furthermore, follow-up by the EPA is needed to clearly articulate the criteria for determining that models are reasonable as well as providing transparent definitions for frequently used terms such as “too steep,” “unstable,” “problematic,” and “credible” (p. 4-38). The SAB recommends assigning weight to certain types of models based on a modified combination of biologic plausibility and statistical considerations, and using somewhat different considerations for comparing AICs than those currently employed in the draft assessment.

Regarding statistical considerations about various models, the SAB recommends a different set of emphases in the priorities for the most reasonable models and gives guidance on the preference for their ordering. First, priority should be given to regression models that directly use individual-level exposure data. Because the NIOSH cohort has rich individual-level exposure data, linear regression of the categorical results should be de-emphasized in favor of models that directly fit individual-level exposure data. Second, among models fit to individual-level exposure data, models that are more tuned to local behavior in the data should be relied on more heavily. Thus, spline models should be given higher priority over transformations of the exposure. Third, the principle of parsimony (the desire to explain phenomena using fewer parameters) should be considered. Attention to this principle becomes even more important as the information in the analysis dataset becomes even more limited. Thus, models with very few estimated parameters should be favored in cases where there are only a few events in the dataset. To elaborate further, in some settings the principle of parsimony may suggest that the most informative analysis will rely upon fixing some parameters rather than estimating them from the data. The impact of the fixed parameter choices can be evaluated in sensitivity analyses. In the draft assessment, fixing the knot when estimating linear spline model fits from relative risk regressions is one such example. Use of AIC can assist with adhering to this principle of parsimony, but its application cannot be used naïvely and without also including scientific considerations. (See further discussion below.) Beyond these recommendations for choosing among models, one advantage of fitting and examining a wide range of models is to get a better understanding of the behavior of the data in the exposure regions of interest. For instance, the models shown in Table 4-13 and Figures 4-5 and 4-6 can be compared, ideally with one or more of these presentations augmented with a few more model fits, including the square root transformation of cumulative exposure, linear regression of categorical results given more categories, and several additional 2-piece linear spline models with different knots. From the comparisons, it is clear that these data suggest a general pattern of the risk rising very rapidly for low-dose exposures and then continuing to rise much more slowly for higher exposures. It is reassuring to observe that many of the fitted models reflect this pattern even though they have different sensitivity to local data.

Results of statistical analyses do not always conform to an a priori understanding of biologic plausibility. When this is the case, investigators need to reassess whether the data are correct, a different approach to model fitting should be employed, or whether the prevailing notion of biologic plausibility should be re-examined. When sufficient exploration of the fitted models has been conducted and a range of models with different properties all suggest a dose-response relationship that would not have been predicted in advance (as is the case in these NIOSH data analyses), then the remaining two considerations should be reviewed. The response to Charge Question 4 further discusses uncertainty in the exposure data. The SAB also encourages finding opportunities to use other evidence from the literature to support the observed dose-response relationship. Specifically, the SAB encourages a discussion of the Swedish sterilization workers study results using the internal comparison group.

The application of AIC for selecting models is acceptable within some constraints as outlined in the following discussion. Burnham and Anderson (2004) is an additional reference that discusses the use of AIC for model selection. (The following discussion is intended to be fairly comprehensive and thus covers points that the SAB did not identify as problematic in the draft assessment.) AIC is an appropriate tool to use for model selection for both nested and non-nested models, provided these models use the same likelihood formulation and the same data. AIC is not the preferred way to characterize model fit. For model selection, (1) AIC is not an appropriate tool for comparing across different models that are fit using different measures, such as comparing a Poisson vs. least squares fit to count data; (2) one should not use AICs to compare models using different transformations of the outcome variable; and (3) comparing AICs from models estimated using different software tools, including different implementations within the same statistical package can be challenging because many calculations of AIC remove constants in the likelihood from the estimated AIC. These AIC features require that users interested in comparing AICs across different software routines (even those within one statistical package) understand exactly what likelihood is being maximized and how the AIC is calculated. AIC can be used to compare the same regression model with the same outcome variable and different predictors whether or not these models are nested. This gives a consistent estimate of the mean-squared prediction error (MSPE), which is one criterion for choosing a model. Finally, the theory behind this MSPE criterion can break down with a large number of models. Thus, naïve applications of AIC for model selection can be problematic (but are not necessarily so in any particular application). In particular, differences in AICs could be an artifact of how the calculation was done. This is a possible difference between the linear and exponential relative risk models applied to the breast cancer incidence data. Although the EPA provided some clarification about its approach in its February 19, 2015 memo to the SAB, the SAB still does not have sufficient information to determine whether or not this is the case.

In conclusion, although the SAB concurs with the EPA's selected model, it believes that aspects of EPA's approach to model selection can be refined and that more transparency in the presentation is needed.

Summary of recommendations:

- Revise the discussion to provide more clarity and transparency as well as making the disposition easier to follow.
- Discarding a model because the fitted curve is "too steep" is only acceptable when there is scientific justification.
- Clearly articulate the criteria for determining that models are reasonable as well as providing transparent definitions for frequently used terms such as "too steep," "unstable," "problematic," and "credible".
- Assign weight to various models based on a modified combination of biological plausibility and statistical considerations; use somewhat different considerations for comparing AICs than those currently employed in the draft assessment.
- Use a different set of emphases in the priorities for the most reasonable models; detailed suggestions are provided by the SAB in this response.

> *2c: For analyses using a two-piece spline model, please comment on whether the method used to identify knots (Section 4.1.2.3 and Appendix D) is transparently described and scientifically appropriate.*
>
> The method used to identify the knots involves a sequential search over a range of plausible knots to identify the value at which the likelihood is maximized. This is scientifically appropriate and a practical solution that is transparently described.

The quote from EPA states "[The knot is preselected and is not considered a parameter in these analyses, consistent with the SAB's concept of parsimony (SAB, 2015)]." However, EPA also states on footnote a to Table D-27 "knots were obtained by doing a grid search by increments of 100 ppm x days and then interpolating where appropriate" and foot note b states "For models with very low knots, alternate knots were obtained from local maximum likelihoods because of the small number of cases informing the slope of the low-exposure spline for low knots (see Figure D-14)." EPA further states on page D-41 (emphasis added) "For the two-piece log-linear model, the single knot was chosen at 100 ppm-days based on a comparison of likelihoods assessed every 100 ppm-day from 100 to 15,000. The best likelihood was at 100 ppm-days. Figure D-15 below shows the likelihood versus the knots. Figure D-15 also suggests a local maximum likelihood near 1,600 ppm-days."

In summary, EPA's description of how the knots for the linear spline and log-linear spline models were found clearly indicate that the knots were not fixed parameters, but rather were optimized numerically and in this way were estimated from the data that were fit. That is, the knots used by EPA for the linear and log-linear spline models were determined using the NIOSH data, so that the knot maximized the likelihood of the spline model. The knots, therefore, were not fixed parameters independent of the NIOSH data, as would be the case in SAB discussion of an example. EPA contradicts itself when it states "[The knot is preselected and is not considered a parameter in these analyses, consistent with the SAB's concept of parsimony (SAB, 2015)].[14]" The latter EPA statement is simply false, because each knot value derived by EPA was in fact optimized (i.e., estimated) by EPA to best fit a corresponding model to a specific set of data. This fact has no relevance at all to the concept of parsimony in model selection, which refers to preference for selecting among different models the one(s) that has (have) the fewest total number (k) of estimated parameters. The parsimony concept is also expressed in the definition of the Akaike Information Criterion (AIC), which is proportional to the value of k, insofar as superior models are identified as those with smaller associated values of AIC. Likewise, a p-value for goodness of model fit is typically evaluated in relation to a corresponding value of the total number of degrees of freedom (DF) associated with that fit, and the latter number is always defined as the total number (n) of data points modeled minus the total number (k) of estimated model parameters, i.e., DF = $n–k$. An invalid reduction in k (e.g., by improperly considering a parameter "fixed" when in fact it was estimated to get a best fit for that model), therefore always improperly inflates the value of DF, which results in an erroneously high p-value for goodness-of-fit that falsely magnifies the likelihood that deviations between data and a model fit to those data are due only to chance (i.e., due only to sampling error).

Misinterpretation of Degrees of Freedom Results in Miscalculated p-values, AIC and Incorrect Model Selection

The "log-linear spline model with knot at 1,600 ppm-days" has three parameters that each were estimated: slope below the knot, slope above the knot, and the knot itself. However, when EPA calculated a corresponding p-value associated with its reported chi-square test for improved fit relative to an associated null model, EPA used only two degrees of freedom for this calculation. This resulted in artificially and erroneously inflating the measure of improved fit used to compare the linear spline model to other models for which p-values were calculated using degrees of freedom that accurately reflected the total number of estimated parameters associated with other model fits being compared. Specifically, EPA did not include the degree of freedom associated with the separate procedure EPA applied to numerically and graphically maximize the log likelihood of each linear spline model for which an optimum knot value was also estimated. By failing to account for the degree of freedom associated with knot-estimation, the p-value EPA reported for each such linear spline model was miscalculated to yield a lower p-value (indicating an unrealistically improved fit) than would be produced had the correct number of degrees of freedom been used by EPA for each such calculation.

In using the approach EPA took in this regard, EPA may have misinterpreted comments of the EPA (2015) Science Advisory Board (SAB) review of the EPA (2014) draft IRIS document, which on pages 12–14 state that:

> the principle of parsimony (the desire to explain phenomena using fewer parameters) should be considered. Attention to this principle becomes even more important as the information in the analysis dataset becomes even more limited. Thus, models with very few estimated parameters should be favored in cases where there are only a few events in the dataset. To elaborate further, in some settings the principle of parsimony may suggest that the most informative analysis will rely upon fixing some parameters rather than estimating them from the data. The impact of the fixed parameter choices can be evaluated in sensitivity analyses. In the draft assessment, **fixing the knot when estimating linear spline model fits from relative risk regressions is one such example**. … differences in AICs could be an artifact of how the calculation was done.

Importantly (as shown above), although the SAB indicated that fixing a knot value can be done as part of a practical approach to knot-value estimation, it also stated that "differences in AICs could be an artifact of how the calculation was done." The SAB unfortunately failed to emphasize (but must be assumed to agree with the fact) that **differences in p-values from chi-square tests of improved fit relative to the null model can also reflect non-meaningful**

artifacts if associated p-value calculations are not done correctly. Specifically, it is not meaningful to compare (as EPA did) a p-value from a Cox linear-regression model of Log(RR) on ppm-days of exposure (defined to be associated with one degrees of freedom for each of the estimated slope of the line) to a p-value from EPA's linear spline model fit (assumed to be associated with only two degrees of freedom corresponding to its two estimated slopes) conditional on a knot value that EPA estimated by minimizing log likelihood in relation to the knot value. EPA incorrectly assumed its optimized knot-value estimate is not associated with one additional degree of freedom. Thus, EPA erroneously deflated the total degrees of freedom associated with their three-parameter linear model by evaluating it as if it had only two degrees of freedom (parameters) associated with it. Consequently, EPA miscalculated the p-value for its spline model resulting in an erroneously low p-values of ~0.07 (see Table 2), when (as explained in more detail in the next section) the correctly calculated p-value is ~2-fold greater (i.e., 0.14 to 0.15) and do not differ meaningfully from p-values associated with the more parsimonious linear Cox regression model (see corrected Table 4-6 discussed in the next section).

Table 2. SAS results given for this model in Table D-33 in Appendix D of EO IRIS 2016

Table D-33. Results of two-piece log-linear spline model for lymphoid cancer mortality, men and women combined, knot at 1,600 ppm-days

Model fit statistics						
Criterion	**Without covariates**	**With covariates**				
-2 LOG L	463.912	**458.640**				
AIC	463.912	**462.640**				
SBC	463.912	466.581				
Testing global null hypothesis: BETA = 0						
Criterion	**Without covariates**	**With covariates**				
Likelihood ratio	5.2722	**2**	**0.0716**			
Score	5.2666	2	0.0718			
Wald	5.1436	2	0.0764			
Analysis of maximum likelihood estimates						
Parameter	**DF**	**Parameter estimate**	**Standard error**	**χ2**	**Pr > ChiSq**	**Hazard ratio**
LIN_0	1	0.0004893	0.0002554	3.6713	0.0554	1.000
LIN_1	1	0.0004864	0.0002563	3.6014	0.0577	1.000

Miscalculated p-values: Example using the log-linear spline model with knot at 1,600 ppm-days"

The likelihood ratio test is used to test whether a fitted model significantly improves the fit of the data by estimating parameters instead of just assuming a baseline (null) model for the data. The likelihood ratio test is evaluated by comparing the likelihood of the model with the estimated parameters and the likelihood of the null model. If the likelihood of the model with the estimated parameters is equal to the likelihood of the null model, then the natural logarithm of the ratio of these likelihoods multiplied by two follow a Chi-Square distribution with as many degrees of freedom as the number of parameters estimated for the fitted model. Thus, if the fit of the baseline (null) model and the model with estimated parameters are not different,

$$Chi - Square(k) = \chi_k^2 = -2\ln\left(\frac{likelihood\ for\ null\ model}{likelihood\ for\ fitted\ model}\right)$$

This can also be written as follows,

$$\chi_k^2 = -2LogL(null\ model) - 2LogL(fitted\ model)$$

Here k is the number of degrees of freedom (k is the number of parameters that were estimated in excess of the parameters estimated for the null model).

For the model in Table 2 (Table D-33 in EO IRIS 2016) the χ_k^2 value was equal to 5.2722 and k was set to 2. This resulted in a p-value of 0.0716. That is, the fitted model was assumed to have two parameters; namely, the slope below the knot and the slope above the knot. The results in Table 2 are from a SAS output for the model specified. The model specified included a knot. This knot was determined so that the likelihood of the spline model was maximized. That is, the knot is another parameter that was searched for outside SAS. Because the estimation of the "knot" was done outside SAS, the SAS program did not count the knot as a parameter and, consequently, the Chi-Square test SAS reported does not reflect the fact that the knot was also estimated. The correct Chi-Square that accounts for the fact that the knot was estimated outside SAS should then be 5.2722, but *k* (the degrees of freedom) should be 3. This corrected calculation would result in a p-value of 0.1529. That is, the corrected p-value indicates that the likelihood of the "log-linear spline model with knot at 1,600 ppm × days" is not different from the likelihood of the null model. In plain words, there is not enough evidence indicating that the fitted log-linear spline model explains the variability in the data any better than the null model.

Miscalculated AICs: Example using the log-linear spline model with knot at 1,600 ppm-days

The Akaike Information Criterion (AIC) is equal to $2k - 2LogL$ where k is the number of parameters estimated for the model and LogL is the logarithm of the likelihood. Here, Table 2 (Table D-33 in EO IRIS 2016) lists the -2LogL as 458.640 and the AIC as 462.640. That is;

$$462.640 = 2k + 458.640$$

The AIC and –2LogL implies that k equals 2. That is, the spline model was assumed to have estimated two parameters; namely, the slope below the knot and the slope above the knot. The results in the Table 2 consist of SAS output for the spline model specified. The model specified included a knot. This knot was pre-assigned (i.e., previously estimated using a separate optimization procedure outside the SAS run), so the likelihood of the model was maximized only *conditional on* the estimated knot-value used for that calculation. Consequently, the knot must be treated as an additional parameter that was estimated outside SAS. Because the estimation of the “knot” was done outside SAS, the SAS run performed by EPA did not count the knot as a model parameter and, consequently, the resulting AIC value it obtained does not reflect that the knot was in fact estimated. EPA could have requested SAS to account properly for the extra degree of freedom properly associated with its estimated knot value, but EPA evidently elected not to make this request of SAS.

The correct AIC, which accounts for the fact that the knot was estimated outside SAS, should instead be

$$\mathrm{AIC} = 464.640 = 2\times 3 + 458.640$$

These differences are summarized in corrected Table 3 below.

Model selection with correct AIC and p-values

EPA selects the “linear spline model with knot at 1,600 ppm × days” for lymphoid for the following reasons:

a) Adequate statistical fit. EPA’s uses the erroneous p-value of 0.07 (Table 1) to select the model arguing that it is close to 0.05. However, the corrected p-value is 0.14 (Table 3) once the fact that the knot was also estimated is accounted for by adding one more degree of freedom to the chi-square distribution. The corrected p-value is now in the range of the p-values for the log-linear and linear models; in fact, it is larger than the p-value (0.13) for the linear model.

b) Adequate visual fit. EPA’s visual fit is dismissed in the footnote of Figure 4-3 of the EO IRIS 2016 report. The footnote reads “(Note that, with the exception of the categorical results and the linear regression of the categorical results, the different models have different implicitly estimated baseline risks; thus, they are not strictly comparable to each other in terms of RR values, i.e., along the y-axis. They are, however, comparable in terms of general shape.)” In addition to the visual-fit caveat listed by EPA in the IRIS report, they failed to indicate that the models are **not** fit to the five nonparametric rate ratios shown in the figure, but rather to the individual cases that includes nine cases of lag-15 EO unexposed workers and 44 cases with lag-15 EO cumulative exposure. That is, the graph shown in Figure 4-3 of the EO IRIS 2016 report does not show all the variability in the full data and visual comparisons can be misleading. Furthermore, the categorical rate ratios are not “the data”, but rather, non-parametric estimate of the rate ratios.

c) Including local fit (visual) **to low-exposure range; linear model.** When the models are plotted against the non-parametric rate ratios of the 44 exposed cases, all models seem to fit the non-parametric models about the same; which is consistent with the calculated p-values and AIC values.

d) AIC within two units of lowest AIC of models considered. EPA's uses the erroneous AIC value of 462.1 to select the model arguing that it is within two units from the lowest AIC (460.2 for the "linear model with log cumulative exposure"). However, the corrected AIC is 464.5 once the fact that the knot was also estimated is accounted for by adding one more parameter in the calculation of the AIC. The corrected AIC for the "linear spline model with knot at 1,600 ppm-days" is now larger than the AIC values for the linear model (463.6) and for the log-linear model (464.4).

Once the errors indicated above concerning calculating p-values, calculating AIC values, and associated adjustments for different calculations of likelihood values are all corrected, EPA's best model for lymphoid should be reconsidered. Using the criteria EPA EO IRIS uses to select a model, the best models for the lymphoid data are the "linear model" followed by the "log-linear model."

Table 3. The following table has been extracted from EO IRIS 2016 Table 4-6 and the p-values and AIC values have been corrected to reflect the degree of freedom for the knot in the spline models and to reflect the likelihood difference between SAS procedures used for linear and log-linear models

Table 4–6. Models considered for modeling the exposure-response data for lymphoid cancer mortality in both sexes in the National Institute for Occupational Safety and Health cohort for the derivation of unit risk estimates

Model[a]	***p-value***[b]	**AIC**[c]	**Comments**
Two-piece spline models			
Linear spline model with knot at 1,600 ppm × days	0.14	464.5	**SELECTED.** Adequate statistical and visual fit, including local fit to low-exposure range; linear model; AIC within two units of lowest AIC of models considered.
Linear spline model with knot at 100 ppm × days	0.11	463.8	Good overall statistical fit and lowest AIC of two-piece spline models, but poor local fit to the low-exposure region, with no cases below the knot.
Log-linear spline model with knot at 1,600 ppm × days	0.15	464.6	Linear model preferred to log-linear (see text above).
Log-linear spline model with knot at 100 ppm × days	0.11	463.8	Good overall statistical fit and tied for lowest AIC_c of two-piece spline models, but poor local fit to the low-exposure region, with no cases below the knot.
Linear (ERR) models (RR = 1 + β × exposure)			
Linear model	0.13	463.6	Not statistically significant overall fit and poor visual fit.
Linear model with log cumulative exposure	0.02	460.6	Good overall statistical fit, but poor local fit to the low-exposure region.

Appendix 1

Model[a]	***p-value***[b]	**AIC**[c]	**Comments**
Linear model with square-root transformation of cumulative exposure	0.053	462.2	Borderline statistical fit, but poor local fit to the low-exposure region.
Log-linear (Cox regression) models (RR = $e^{\beta \times \text{exposure}}$)			
Log-linear model (standard Cox regression model)	0.22	464.4	Not statistically significant overall fit and poor visual fit.
Log-linear model with log cumulative exposure	0.02	460.4	Good overall statistical fit; lowest AIC[c] of models considered; low-exposure slope becomes increasingly steep as exposures decrease, and large unit risk estimates can result; preference given to the two-piece spline models because they have a better ability to provide a good local fit to the low-exposure range.
Log-linear model with square-root transformation of cumulative exposure	0.08	462.8	Not statistically significant overall fit and poor visual fit.

[a] All with cumulative exposure as the exposure variable, except where noted, and with a 15-yr lag.

[b] *p*-values from likelihood ratio test, except for linear regression of categorical results, where Wald *p*-values are reported. $p < 0.05$ considered "good" statistical fit; $0.05 < p < 0.10$ considered "adequate" statistical fit if significant exposure-response relationships have already been established with similar models.

[c] AICs for linear models are directly comparable and AICs for log-linear models are directly comparable. However, for the lymphoid cancer data, SAS proc NLP (where NLP = nonlinear programming) consistently yielded −2LLs and AICs about 0.4 units lower than proc PHREG for the same models, including the null model, presumably for computational processing reasons, and proc NLP was used for the linear RR models. Thus, AICs for linear models are equivalent to AICs about 0.4 units higher for log-linear models. No AIC was calculated for the linear regression of categorical results.

Note: In order to make the AICs comparable for different models, the AIC's for the linear models have been increased by 0.4 to reflect the discrepancy in the -2LogL values reported by the SAS proc NLP and by SAS PHREG (as indicated in green in this table).

Figures 1 to 4 are versions of EPA's Figure 4-3. A model (TrueLogL – dotted light blue line in the graphs) was added to relieve the caveat posed by EPA in the footnote to Figure 4-3 about the visual comparability of fitted models. The TrueLogl model is an approximation to the correct visual representation of the log-linear (standard Proportional Hazards Model fit to the NIOSH full data set) after adjusting for the difference in baseline risks between the rate ratios and the loglinear model. In Figures 1 to 4, all the individual RR (categorical) in the light blue box of the figure are summarized by the red dot in the light blue box (EPA's 5 RRs for the last quartile). Similarly, all the individual RR (categorical) in the light yellow box of the figure are summarized by the red dot in the light yellow box (EPA's 5 RRs for the third quartile). In the same way, all the individual RR (categorical) in the light green box of the figure are summarized by the red dot in the light green box (EPA's 5 RRs for the second quartile). Finally, all the individual RR (categorical) in the clear box, next to the vertical axis of the figure, are summarized by the red dot in the clear box (EPA's 5 RRs for the first quartile).

Appendix 1

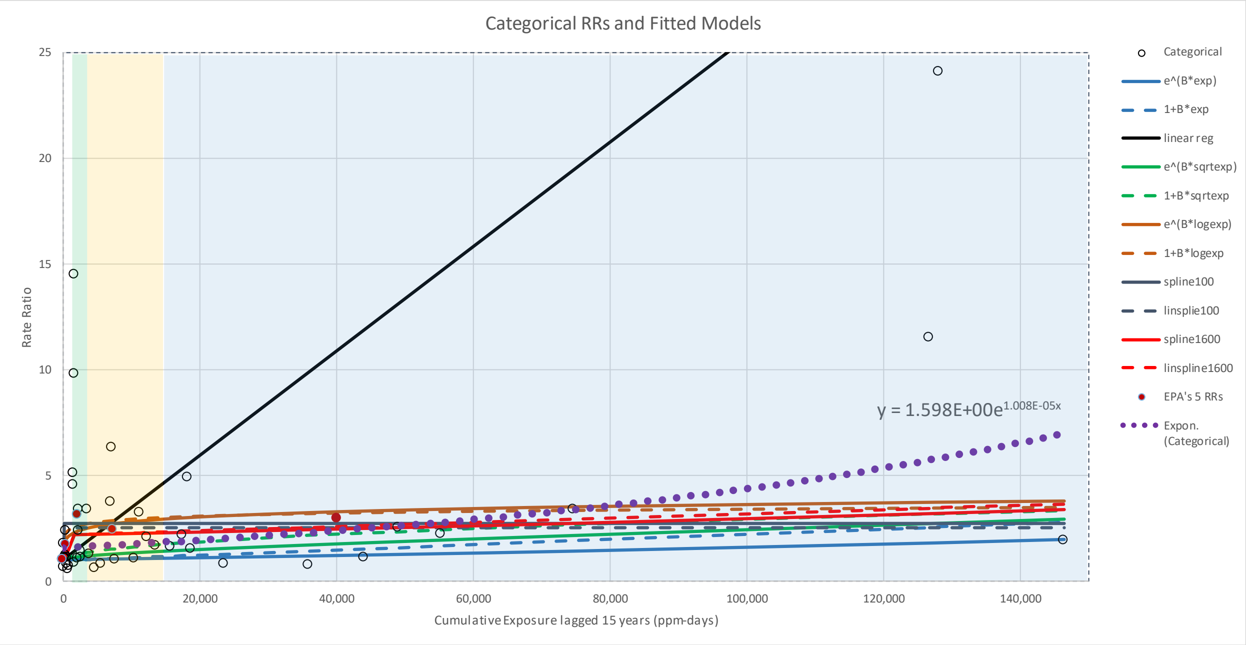
Figure 1 shows all EPA models plotted versus the individual nonparametric rate ratios (categorical) and grouped rate ratios (EPA's 5 RRs). The range of cumulative exposures when the rate ratios for all cases are plotted is much bigger than the range of cumulative exposures when the rate ratios are averaged over several cases (EPA's 5 RRs). The variability of the rate ratios for the individual cases (categorical) is much larger than the variability of the rate ratios averaged over several cases (EPA's 5 RRs). Except for the unacceptable linear model fit to four rate ratios (linear reg), all models fit approximately the same in Figure 1. The model Expon. (Categorical) is a plot of the approximate log-linear model (e^(B*exp)) adjusted by dividing the model for the hazard rate by the baseline hazard rate of the nonparametric estimates.

Figure 2 shows an expansion of the low-left corner of Figure 1. These are all EPA models plotted versus the nonparametric rate ratios with values between 0 and 3.5 and cumulative exposures between 0 and 40,000 ppm-days. This graph resembles Figure 4-3 of the EO IRIS 2016 report with the exception that rate ratios based on individual cases (categorical) that are in the range of the graph are plotted in addition to the aggregated four points used by EPA (EPA's 5 RRs).

Figure 3 is the same as Figure 1 except that the vertical scale is shown using a logarithmic scale of the rate ratios to visualize the linear difference between the fitted models and the rate ratios.

Figure 4 is the same as Figure 2 except that the vertical scale is shown using a logarithmic scale of the rate ratios to visualize the linear difference between the fitted models and the rate ratios.

Figure 1. EPA models plotted against all lymphoid rate ratios in the NIOSH data



Appendix 1

Figure 2. EPA models plotted against all lymphoid rate ratios in the NIOSH data in the low exposure concentration range and with the rate ratio truncated to the same range of EPA's Figure 4-3.

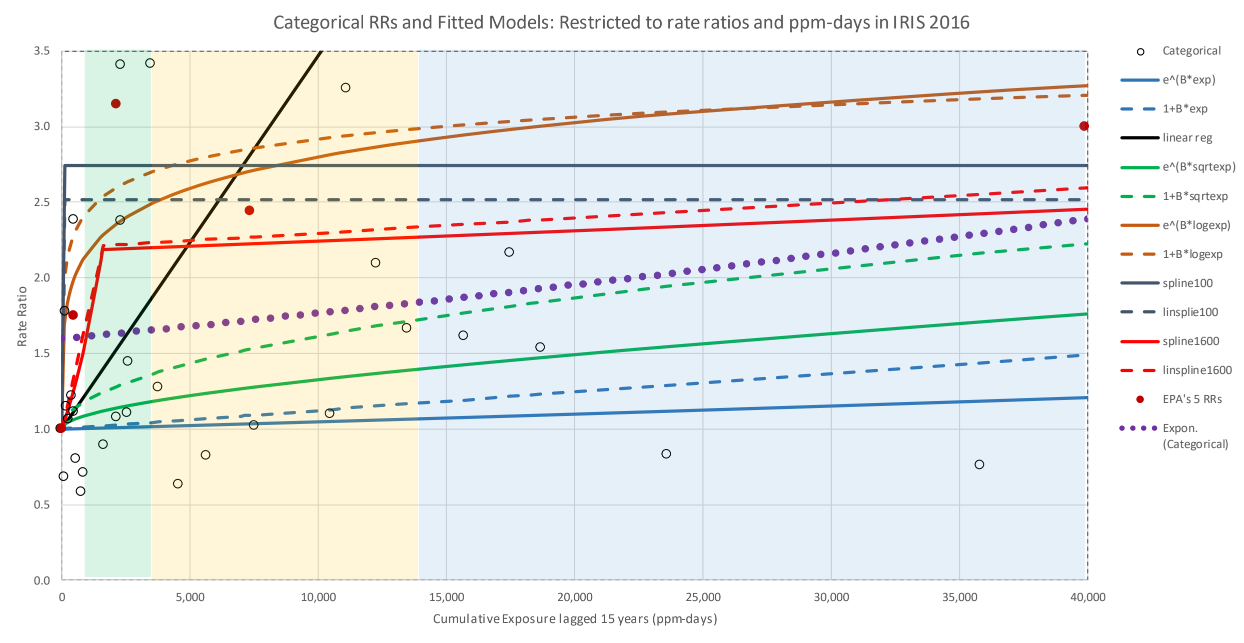
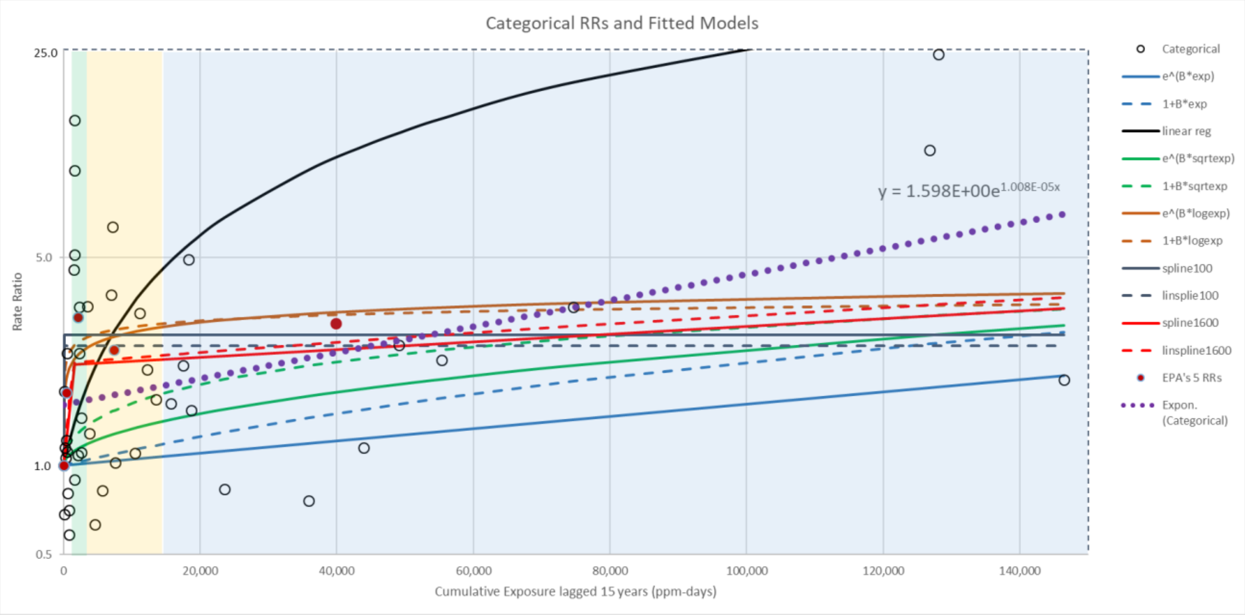
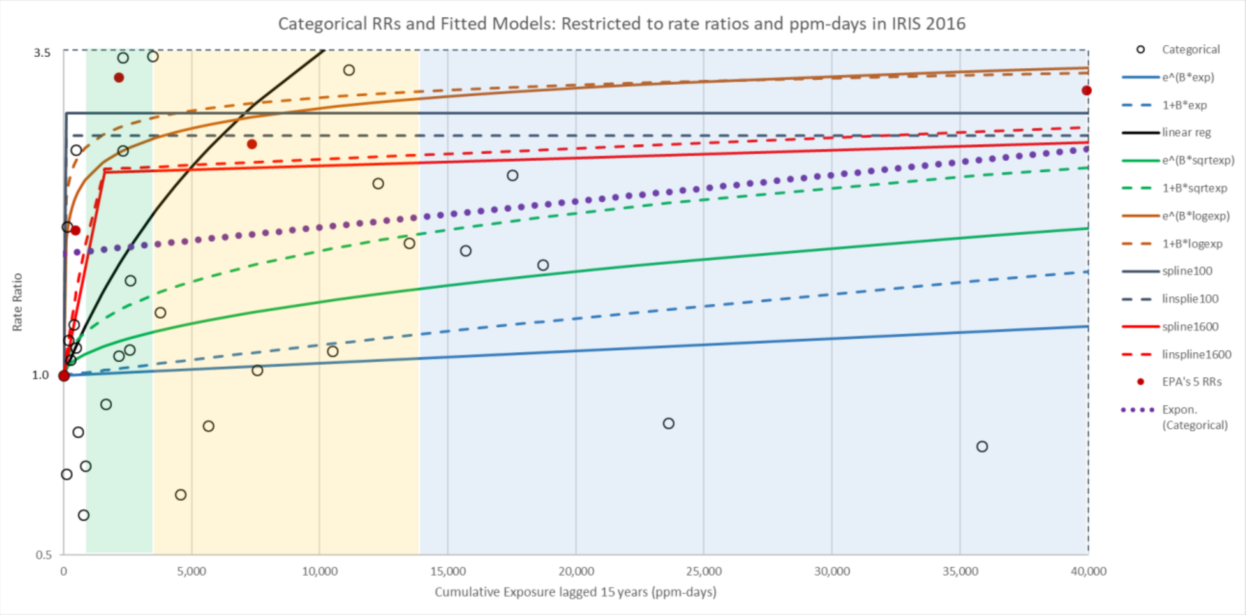


Figure 3. EPA models plotted against the logarithm of all lymphoid rate ratios in the NIOSH data



Appendix 1

Figure 4. EPA models plotted against the logarithm all lymphoid rate ratios in the NIOSH data in the low exposure concentration range and with the rate ratio truncated to the same range of EPA's Figure 4-3.



Appendix 2

Brief Summary of Epidemiological Data for EO

M. Jane Teta, Dr.P.H., M.P.H.
Exponent Health Sciences

The relevant epidemiology, despite the large number of studies published over a forty-year period, are not supportive of a determination that EO is a human carcinogen. While interest has centered on leukemia, other blood related malignancies, and recently breast cancer: (1) there are numerous inconsistencies across the studies, (2) elevated risks above background are found in isolated studies and the effect size is of small magnitude, and (3) there is an absence of a clear exposure-response relation for any specific cancer type.

Examination of the specific cancer subtypes (leukemia, non-Hodgkin's lymphoma [NHL], Hodgkin's disease [HD], multiple myeloma [MM] and lymphohematopoietic cancers [LH] overall) illustrates the absence of clear evidence of carcinogenicity and no clear choice for a target organ should a dose-response be attempted. Table 1 summarizes the individual and overall findings from the EO studies for leukemia. Taking the ratio of the total observed cases and the total expected number of cases yields a summary risk estimate. The total number of deaths due to leukemia is 64 with 56.86 expected for an SMR /SIR of 1.13 (95% CI: 0.87-1.44). It is noteworthy that Hogstedt's increase was mainly attributable to myeloid leukemias, while Steenland focused on lymphocytic leukemia in the lymphoid category. As shown by Shore and Teta in their meta-analyses, Hogstedt is an outlier that is statistically different in findings from the other studies, i.e., a cause of heterogeneity. Furthermore, it is incorrect to include a cluster which gave rise to the hypothesis in a summary risk estimate. Excluding Hogstedt, yields 57 observed leukemias and 56.06 expected for an SMR/SIR of 1.02 (95% CI: 0.77, 1.32). Clearly Hogstedt's hypothesis of EO as a cause of leukemia has not been confirmed.

Appendix 2

Table 1. Leukemia in Epidemiology Studies of Ethylene Oxide

Publication	Observed	Expected	Obs./Exp. (95% CI)
Hogstedt 1979, 1986, 1988	7	0.80	9.21* (3.70, 19.0)
Lymphocyctic	2	---	---
Myeloid	3	---	---
NOS	2	---	---
Hagmar 1991/Hagmar 1995/ Mikoczy 2011	5	3.58	1.40 (0.45, 3.26)
Thiess 1981/Kiesselbach 1990	2	2.35	0.85 (0.10, 3.07)
Morgan 1981/Divine 1990	0	0.60	0.00 (0.00, 6.57)
Greenberg 1990/Teta 1993/ Swaen 2009	11	11.8	0.93 (0.47, 1.67)
Steenland 1991/Stayner 1993/ Steenland 2004	29	29.3	0.99 (0.71, 1.36)
Bisanti 1993	2	0.30	6.50 (0.79, 23.5)
Gardner 1989/Coggon 2004	5	4.60	1.08 (0.35, 2.51)
Olsen 1997	2	3.00	0.67 (0.08, 2.40)
Norman 1995	1	0.54	1.85 (0.05, 10.3)
Summary	64	56.9	1.13 (0.87, 1.44)
Summary (-Hogstedt)	57	56.1	1.02 (0.77, 1.32)

For HD there were 17 observed compared to 10.84 expected (1.57; 95% CI: 0.91-2.51) (Table 2). The Swaen case-control study was included and an expected number was derived to combine these results with those of the cohort studies. (The proportion of controls exposed, 5%, was applied to the case group of 10 cases yielding an expected exposed of 0.5). Relying only on the two strongest studies (Swaen 2009 and Steenland 2004) yields for HD, 6 vs. 6.54 (0.92; 95% CI: 0.34, 2.0). The Swaen 2009 UCC cohort had no deaths due to HD.

Table 2. Hodgkin Disease in Epidemiology Studies of Ethylene Oxide

Publication	Observed	Expected	Obs./Exp. (95% CI)
Hogstedt 1979, 1986, 1988	0	---	---
Hagmar 1991/Hagmar 1995/ Mikoczy 2011	1	1.31	0.76 (0.02, 4.25)
Thiess 1981/Kiesselbach 1990	---	---	---
Morgan 1981/Divine 1990	3	0.40	8.34* (1.68, 24.4)
Greenberg 1990/Teta 1993/ Swaen 2009	0	1.70	0.00* (0.00, 0.22)
Steenland 1991/Stayner 1993/ Steenland 2004	6	4.84	1.24 (0.53, 2.43)
Bisanti 1993	---	---	---
Gardner 1989/Coggon 2004	2	1.05	1.91 (0.23, 6.89)
Olsen 1997	2	0.70	2.86 (0.35, 10.3)
Norman 1995	0	0.34	0.00 (0.00, 10.9)
Swaen 1996	3	0.50	8.50* (1.40, 39.9)
Summary	17	10.8	1.57 (0.91, 2.51)

Two studies provided no data for MM (Kiesselbach 1990 and Bisanti 1993) and four others failed to provide expected values (Hogstedt 1988, Divine 1990, Olsen 1997, and Swaen 2009) (Table 3). Upon contacting Dow, we were able to obtain the expected number of 5.1 for MM. Based on the studies with complete information, there are 22 observed and 24.0 expected for a summary estimate of 0.92 (Table 3). This result is heavily weighted by the largest study, Steenland et al. 2004, who reported 13 cases vs. 14.13 expected (SMR= 0.92). This summary risk estimate does not indicate an association with MM.

Table 3. Multiple Myeloma in Epidemiology Studies of Ethylene Oxide

Publication	Observed	Expected	Obs./Exp. (95% CI)
Hogstedt 1979, 1986, 1988	0	---	---
Hagmar 1991/Hagmar 1995/ Mikoczy 2011	2	2.08	0.96 (0.12, 3.47)
Thiess 1981/Kiesselbach 1990	---	---	---
Morgan 1981/Divine 1990	0	---	---
Greenberg 1990/Teta 1993/ Swaen 2009	3	5.10	0.59 (0.12, 1.72)
Steenland 1991/Stayner 1993/ Steenland 2004	13	14.1	0.92 (0.49, 1.57)
Bisanti 1993	---	---	---
Gardner 1989/Coggon 2004	3	2.50	1.20 (0.25, 3.49)
Olsen 1997	1	NR	NR
Norman 1995	1	0.23	4.34 (0.11, 24.2)
Summary	22	24.0	0.92 (0.57, 1.39)

Using the same method of pooling the observed and expected values of NHL across the different studies results in a meta-SMR/SIR estimate of 1.12 based on 62 observed and 55.4 expected, a small, non-statistically significant increase (Table 4).

Table 4. Non-Hodgkins Lymphoma in Epidemiology Studies of Ethylene Oxide

Publication	Observed	Expected	Obs./Exp. (95% CI)
Hogstedt 1979, 1986, 1988	2	---	---
Hagmar 1991/Hagmar 1995/ Mikoczy 2011	9	6.25	1.44 (0.66, 2.73)
Thiess 1981/Kiesselbach 1990	---	---	---
Morgan 1981/Divine 1990	0	0.90	0.00 (0.00, 4.04)
Greenberg 1990/Teta 1993/ Swaen 2009	12	11.5	1.05 (0.54, 1.83)
Steenland 1991/Stayner 1993/ Steenland 2004	31	31.0	1.00 (0.72, 1.35)
Bisanti 1993	3	0.20	16.9* (3.49, 49.5)
Gardner 1989/Coggon 2004	7	4.80	1.46 (0.59, 3.02)
Olsen 1997	5	NR	NR
Norman 1995	0	0.76	0.00 (0.00, 4.85)
Summary	62	55.4	1.12 (0.86, 1.43)

Examination across the ten studies of all LH cancers yields a non-statistically significant increase based on 175 observed vs. 156.97 expected (Meta-SMR/SIR = 1.11; 95% CI: 0.96, 1.29) (Table 5). Exclusion of Hogstedt would result in a weak excess (1.07) and narrow confidence interval (95% CI: 0.91, 1.25).

Table 5. All Lymphopoietic and Hematopoietic Cancers in Epidemiology Studies of Ethylene Oxide

Publication	Observed	Expected	Obs./Exp. (95% CI)
Hogstedt 1979, 1986, 1988	9	2.00	4.59* (2.10, 8.70)
Hagmar 1991/Hagmar 1995/ Mikoczy 2011	18	14.4	1.25 (0.74, 1.98)
Thiess 1981/Kiesselbach 1990	5	4.99	1.00 (0.32, 2.34)
Morgan 1981/Divine 1990	3	3.00	1.01 (0.20, 2.96)
Greenberg 1990/Teta 1993/ Swaen 2009	27	30.4	0.89 (0.59, 1.29)
Steenland 1991/Stayner 1993/ Steenland 2004	79	79.0	1.00 (0.79, 1.24)
Bisanti 1993	5	0.70	7.00* (2.27, 16.4)
Gardner 1989/Coggon 2004	17	12.9	1.30 (0.77, 2.10)
Olsen 1997	10	7.70	1.29 (0.62, 2.38)
Norman 1995	2	1.88	1.06 (0.13, 3.84)
Summary	175	157.0	1.11 (0.96, 1.29)
Summary (-Hogstedt)	166	155.0	1.07 (0.91, 1.25)

As discussed above, Steenland et al. (2004) grouped three LHC cancers into the “lymphoid” category and reported some positive findings for men only. This category included lymphocytic leukemias only. The original cluster reported by Hogstedt in 1979 consisted of myeloid leukemias (Table 2). The results from the only other study to examine the lymphoid category as defined by NIOSH (UCC cohort) are inconsistent with the NIOSH results (Swaen 2009). From an internal analysis using Cox proportional hazard model, no evidence of an exposure–related response was observed by Swaen et al. using the UCC EO cohort. In fact, the females in the NIOSH study are also inconsistent with the male findings for lymphohematopoietic and “lymphoid” tumors (Steenland 2004).

Steenland et al. also examined both incidence and mortality from breast cancer for the sterilizer cohort (Steenland 2003, 2004). Among the overall results for this disease endpoint among other studies, only Norman et al. (1995) reported an increase (Table 6). Hogstedt enumerated all the cancers from his numerous cohorts and updates. No breast cancer cases were identified. Similarly, there was no excess among the hospital workers studies by Coggon et al. (2004), even among those with “continual” exposure (5 observed, 7.2 expected). The data related to breast cancer derived predominately from the NIOSH studies of sterilant workers with 102 deaths and 103 expected for an SMR of 0.99 (95% CI: 0.81-1.20) (Steenland 2004) and 319 incident cases with 367 expected for a statistically significant deficit of 0.87 (95% CI: 0.77-0.97) (Steenland 2003) due to underascertainment of cases. When examined in various exposure subgroup analyses, however, NIOSH concluded there was some evidence of an increase for breast cancer.

Table 6. Ethylene Oxide Epidemiology Studies of Female Breast Cancer

Study	Observed	Expected	Obs./Exp. (95% CI)
Coggon et al. 2004	11	13.1	0.84 (0.42, 1.51)
Steenland et al. 2004	102	103.0	0.99 (0.81, 1.20)
Steenland et al. 2003	319	367.0	0.87* (0.77, 0.97)
Mikoczy et al. 2011	41	50.9	0.81 (0.58, 1.09)
Norman et al. 1995	12	7.0	1.72 (0.93, 2.93)
Hogstedt et al. 1986	0	---	---
Summary (incident cases only)	372	424.9	0.88* (0.79, 0.97)
Summary (mortality cases only)	113	116.1	0.97 (0.80, 1.17)

EPA recognizes that magnitudes of increased risks for breast cancer were not large and implies that the evidence is weaker than that for lymphoid tumors. Despite these issues, EPA proceeds to introduce breast cancer as a target organ in the IRIS Assessment and inappropriately develops a risk value. Uncertainties described by Steenland et al. (2003) related to the breast cancer incidence study are dismissed as unimportant by EPA. EPA agrees with Steenland that the breast cancer incidence findings are not conclusive, due to inconsistencies in the exposure-response and an incomplete cancer ascertainment. Using these data, the slopes of EPA's attempted exposure-response analyses were non-statistically significant or biologically uninterpretable, leading them to employ novel approaches for quantitative risk assessment. The modeling challenges could be anticipated given Steenland's statement of uncertainty with respect to breast cancer, "The dip in the spline curve in the region of higher exposures suggested an inconsistent or non-monotonic risk with increasing exposure."

The Agency downplays the potential for selection bias based on the consistency in the incidence study between results from full cohort and those from the subgroup interviewed (68% of study subjects). Selection bias (referred to by Steenland as "possible biases due to patterns of non-response") remains a concern, however, with duration reported as a stronger risk factor than cumulative exposure in both analyses. Those who work longer stay in the area longer and are more likely to get picked up in the state tumor registries and be found for interview, therefore with the potential to impact the results of both analyses. Shorter duration workers with lower exposures are more likely to leave the area and not be captured in the overall analyses and less likely to be interviewed. Their diagnoses get missed, creating a possible biased positive exposure-response. Steenland recognized this limitation and admitted he was unable to fully address it and listed it as one of his uncertainties:

> A second possible bias was the preferential ascertainment of breast cancer among women with stable residence in states with cancer registries; women with stable residency might be expected to have longer duration of employment in companies

> under study, and hence greater cumulative exposure. Unfortunately, we did not have residential history, limiting our ability to explore this possibility.

The more recent study by Mikoczy et al. (2011) has been cited as supportive of an association with breast cancer, in spite of an overall deficit (SIR=0.81; 95% CI: 0.58-1.09) based on 41 cases observed. With 15-year latency it is 0.86, also suggesting no increase. Similar to NIOSH, however, the two higher cumulative exposure groups (of three total group) had statistically significant elevated rates of breast cancer (2.76; 95% CI: 1.20-6.33 and 3.55; 95% CI: 1.58-7.93) in an *internal* Poisson analysis, due, however, to a substantial and statistically significant deficit of breast cancer in the low dose reference group (SIR=0.52; 95% CI: 0.25-0.96). There are clearly advantages to comparing workers to workers in epidemiology studies to overcome possible biases in external comparisons to the general population. However, there may also be disadvantages to using an internal comparison group that are not recognized. One danger is selecting a referent group that has an unusual excess or deficit of the disease of interest as illustrated in this study. This illustrates the problem that can arise from internal comparisons and should not always to be preferred despite what EPA contends.

In addition to LH cancers, EPA uses breast cancer as a target endpoint. We conclude that the choice of breast cancer as a target organ for EO dose-response assessment is not justified for several reasons: (1) EPA agrees that the evidence for breast cancer is even weaker than the evidence for the lymphoid category, (2) the NIOSH findings suffer from potential selection biases, show a non-monotonic increase in risk with increasing exposure, and neither mortality nor incidence rates overall exceed background rates in the general population, and (3) the breast cancer findings from the other epidemiology studies are equivocal.

There is no obvious target organ for an EO exposure-response assessment for a quantitative risk assessment. Given the weak epidemiology evidence for carcinogenicity, the lack of consistency or a clear exposure-response, the selection of a specific target organ is problematic. Using cumulative exposure as the exposure metric and the standard proportional hazard modeling, none of the slopes for the endpoints of interest are statistically significant (Valdez-Flores, Sielken, and Teta 2010). Despite the absence of a clear exposure-response for any one of the combinations, the authors proceeded to use EPA's standard procedure for unit risk estimation and estimation of exposure associated with a one-in-a-million risk. This approach was adopted by Scientific Committee on Occupational Exposure Limits (SCOEL) for the European Union in 2012 for occupational standard setting.

References

1. Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide (Revised External Review Draft): United States Environmental Protection Agency: Office of Research and Development August 2014.

2. Bisanti L, Maggini M, Raschetti R, Alegiani SS, Ippolito FM, Caffari B, et al. Cancer mortality in ethylene oxide workers. Br J Ind Med 1993;50(4):317-24.

3. Coggon D, Harris EC, Poole J, Palmer KT. Mortality of workers exposed to ethylene oxide: extended follow up of a British cohort. Occup Environ Med 2004;61(4):358-62.

4. Divine BJ. Update of Texas Morgan Study. Presentation at the American Conference of Occupational Medicine Meeting (unpublished). In. Houston, Texas; 1990.

5. Gardner MJ, Coggon D, Pannett B, Harris EC. Workers exposed to ethylene oxide: a follow up study. Br J Ind Med 1989;46(12):860-5.

6. Greenberg HL, Ott MG, Shore RE. Men assigned to ethylene oxide production or other ethylene oxide related chemical manufacturing: a mortality study. Br J Ind Med 1990;47(4):221-30.

7. Hagmar L, Mikoczy Z, Welinder H. Cancer incidence in Swedish sterilant workers exposed to ethylene oxide. Occup Environ Med 1995;52(3):154-6.

8. Hagmar L, Welinder H, Linden K, Attewell R, Osterman-Golkar S, Tornqvist M. An epidemiological study of cancer risk among workers exposed to ethylene oxide using hemoglobin adducts to validate environmental exposure assessments. Int Arch Occup Environ Health 1991;63(4):271-7.

9. Hogstedt C, Aringer L, Gustavsson A. Epidemiologic support for ethylene oxide as a cancer-causing agent. JAMA 1986;255(12):1575-8.

10. Hogstedt C, Rohlen O, Berndtsson BS, Axelson O, Ehrenberg L. A cohort study of mortality and cancer incidence in ethylene oxide production workers. Br J Ind Med 1979;36(4):276-80.

11. Hogstedt LC. Epidemiological studies on ethylene oxide and cancer: an updating. IARC Sci Publ 1988(89):265-70.

12. Kiesselbach N, Ulm K, Lange HJ, Korallus U. A multicentre mortality study of workers exposed to ethylene oxide. Br J Ind Med 1990;47(3):182-8.

13. Mikoczy Z, Tinnerberg H, Bjork J, Albin M. Cancer incidence and mortality in Swedish sterilant workers exposed to ethylene oxide: updated cohort study findings 1972-2006. Int J Environ Res Public Health 2011;8(6):2009-19.

14. Morgan RW, Claxton KW, Divine BJ, Kaplan SD, Harris VB. Mortality among ethylene oxide workers. J Occup Med 1981;23(11):767-70.

15. Norman SA, Berlin JA, Soper KA, Middendorf BF, Stolley PD. Cancer incidence in a group of workers potentially exposed to ethylene oxide. Int J Epidemiol 1995;24(2):276-84.

16. Olsen GW, Lacy SE, Bodner KM, Chau M, Arceneaux TG, Cartmill JB, et al. Mortality from pancreatic and lymphopoietic cancer among workers in ethylene and propylene chlorohydrin production. Occup Environ Med 1997;54(8):592-8.

17. Shore RE, Gardner MJ, Pannett B. Ethylene oxide: an assessment of the epidemiological evidence on carcinogenicity. Br J Ind Med 1993;50(11):971-97.

18. Stayner L, Steenland K, Greife A, Hornung R, Hayes RB, Nowlin S, et al. Exposure-response analysis of cancer mortality in a cohort of workers exposed to ethylene oxide. Am J Epidemiol 1993;138(10):787-98.

19. Steenland K, Stayner L, Deddens J. Mortality analyses in a cohort of 18 235 ethylene oxide exposed workers: follow up extended from 1987 to 1998. Occup Environ Med 2004;61(1):2-7.

20. Steenland K, Stayner L, Greife A, Halperin W, Hayes R, Hornung R, et al. Mortality among workers exposed to ethylene oxide. N Engl J Med 1991;324(20):1402-7.

21. Steenland K, Whelan E, Deddens J, Stayner L, Ward E. Ethylene oxide and breast cancer incidence in a cohort study of 7576 women (United States). Cancer Causes Control 2003;14(6):531-9.

22. Swaen GM, Burns C, Teta JM, Bodner K, Keenan D, Bodnar CM. Mortality study update of ethylene oxide workers in chemical manufacturing: a 15 year update. J Occup Environ Med 2009;51(6):714-23.

23. Swaen GM, Slangen JM, Ott MG, Kusters E, Van Den Langenbergh G, Arends JW, et al. Investigation of a cluster of ten cases of Hodgkin's disease in an occupational setting. Int Arch Occup Environ Health 1996;68(4):224-8.

24. Teta MJ, Benson LO, Vitale JN. Mortality study of ethylene oxide workers in chemical manufacturing: a 10 year update. Br J Ind Med 1993;50(8):704-9.

25. Teta MJ, Sielken RL, Jr., Valdez-Flores C. Ethylene oxide cancer risk assessment based on epidemiological data: application of revised regulatory guidelines. Risk Anal 1999;19(6):1135-55.

26. Thiess AM, Schwegler H, Fleig I, Stocker WG. Mutagenicity study of workers exposed to alkylene oxides (ethylene oxide/propylene oxide) and derivatives. J Occup Med 1981;23(5):343-7.

27. Valdez-Flores C, Sielken RL, Jr., Teta MJ. Quantitative cancer risk assessment based on NIOSH and UCC epidemiological data for workers exposed to ethylene oxide. Regul Toxicol Pharmacol 2010;56(3):312-20.