Comments to the National Academies of Sciences, Engineering, and Medicine (NASEM) Committee reviewing of Texas Commission on Environmental Quality's Ethylene Oxide Development Support Document (TCEQ DSD)

July 19, 2024

Kenneth Bogen, [1] Dr.P.H., DABT, James Bus, [2] Ph.D., DABT, ATS, Bhaskar Gollapudi, [3] Ph.D., Christopher Kirman, [4] M.S., Abby Li,[2] Ph.D., Richard Reiss,[2] Sc.D., Patrick Sheehan,[2] Ph.D., Steave Su,[2] M.P.H., Jane Teta,[2] Dr.P.H., M.P.H. [5]

The National Academies of Sciences, Engineering, and Medicine (NASEM) committee's scientific review of the Texas Commission on Environmental Quality (TCEQ) 2020 Ethylene Oxide (EtO) Carcinogenic Dose-Response Assessment Development Support Document (DSD) is important because TCEQ's cancer slope is approximately 2000-fold lower than EPA's cancer slope. The major reason for this difference is in the selection of the dose-response model that is applied to the NIOSH cohort. Another difference in the quantitative risk assessment is that EPA combines risks for both lymphoid mortality and breast cancer incidence (with lymphoid contributing approximately 87% of the total EPA IRIS cancer risk estimate [6]), while TCEQ only includes lymphoid mortality. Appropriately, NASEM invited the U.S. Environmental Protection Agency (EPA) to present their approach (EPA, 2024a) on June 27, 2024 (Meeting 4).

Unfortunately, EPA's presentation contained inaccurate, incomplete, and misleading information. For example, EPA presented figures from the IRIS assessment that EPA alleges shows a superior "visual fit" to its model compared to the TCEQ model indicating that the TCEQ model underpredicts the data (e.g. EPA slide 15 & 17). However, EPA omitted the crucial EPA footnote included in every EPA IRIS figure of relative risk (RR) that warns against such comparisons along the y-axis. EPA's reference slides include factual errors, errors of omission and logical fallacies in comparing categorical and continuous modeling to discount TCEQ's correct explanation for why comparisons of RR cannot be made. These slides incorrectly claim that TCEQ's graphs are of RR despite TCEQ's careful labeling of the y-axis and detailed figure legend explanations to the contrary. Ironically, TCEQ graphically illustrates why EPA IRIS footnotes correctly warn against the type of comparisons EPA relies on to select their model.

[1] Versar Global Solutions through Exponent, Inc, Oakland, CA
[2] Exponent, Inc, Oakland, CA; NY, NY; Alexandria, VA
[3] Toxicology Consultant, Midland, MI
[4] SciPinion, Bozeman, MT
[5] Funding sources for the authors include American Chemistry Council Ethylene Oxide Panel, the Ethylene Oxide Task Force or the Ethylene Oxide Sterilization Association. However, the analysis and opinions included in this white paper are solely the authors' and do not necessarily reflect those of the funding sources
[6] EPA IRIS (2016), p.4-74

EPA also incorrectly stated that EPA SAB instructed EPA to apply the 2-piece spline model to lymphoid cancers. EPA SAB (2015) rejected the draft IRIS selection of a linear regression of the categorical "grouped" estimates, but never recommended application of the 2-piece spline to the lymphoid cancers. EPA IRIS essentially ignored the essence of SAB's advice by relying on misleading visual fit comparisons with the categorical "grouped" estimates that EPA SAB had rejected as the basis for EPA modeling. It is notable that the final 2-piece spline model selected for lymphoid cancer was never peer reviewed by EPA SAB.

EPA SAB (2015) emphasized that **"any model that is to be considered reasonable for risk assessment must have a dose-response form that is both biologically plausible and consistent with the observed data."** The TCEQ (2020) DSD more faithfully follows EPA SAB (2015) advice because it emphasizes the integration of the biological and epidemiological evidence and sound statistical principles when considering which type of Cox proportional hazards (CPH) model to apply to the NIOSH lymphoid cancer mortality and whether or not to include breast cancer incidence. This is in direct contrast with the EPA IRIS (2016) exposure-response assessment, which is based on an extensive statistical modeling exercise absent any consideration of the epidemiological weight of evidence and biological plausibility. While the IRIS assessment includes summaries of the genotoxicity, toxicology, epidemiology and toxicokinetics for hazard assessment, there is virtually no integration of these important lines of evidence applied to the final quantitative cancer risk assessment.

The EPA IRIS statistical modeling of the mortality data builds on a backdrop of a large number (>50) of statistical models that were applied by Steenland et al. (2004) [7]. The only continuous dose-response models that were reported by Steenland et al. (2004) to have statistically significant positive slopes were

(1) Lymphohematopoietic (LH) cancer mortality: males, log cumulative exposure model with 15 yr lag
(2) Lymphoid cancer mortality (a subset of LH): males, log cumulative exposure model, and 15 yr lag
(3) Breast cancer mortality: log cumulative exposure model and 20 year lag

If Steenland et al. (2004) had accounted for the lag in the analysis, none of these 3 models would have been statistically significantly different from a model with no exposure response (the null model). EPA contracted with Dr. Steenland to expand the statistical analyses beyond the original published approaches to include a large number of additional models including 2-

[7] Steenland et al. (2004) modeled the data using a variety of exposure metrics (duration, average, maximum, cumulative, and log cumulative exposure) and both continuous and categorical exposure variables. For each of these combinations, a variety of lags were also examined (no lag, 5, 10, 15 and 20).

slope spline models. The EPA (2005) carcinogen risk assessment guidelines caution against applying multiple curve-fitting models:

> "Another problem occurs when a multitude of alternatives are presented without sufficient context to make a reasoned judgment about the alternatives. This form of model uncertainty reflects primarily the availability of different computer models and not biological information about the agent being assessed or about carcinogenesis in general."

EPA's response to the NASEM committee member questions (presented below as Questions 1 and 2) illustrates this over-reliance on incorrect statistics and misleading visual fit comparisons.

NASEM Committee Member Question 1: How was local fit determined?

EPA responded by presenting slide 28 (copied below as Table 1), but never explained that for log linear cumulative exposure model (TCEQ's approach), the "poor local fit to all categorical points" is based solely on an incorrect visual fit criteria. TCEQ (2020) correctly explains that EPA's visual fit is based on invalid subjective comparisons between the model and grouped categorical estimates along the y-axis. The TCEQ log-linear CPH model statistically

Table 1. EPA (2024; PowerPoint Presentation (nationalacademies.org)) Reference Slide 28 presented to NASEM on June 27, 2024:

Models for lymphoid cancer mortality. Consolidated results from EtO IRIS Assessment

model name	-2*log likelihood (deviance)	AIC	p-value	local low dose fit	low dose shape	Plateau-like shape	Unit risk/ppm (low dose)	Considerations
Two-piece linear spline (knot 1600 ppm-days)	458.1	462.1*	0.07*	yes	linear	yes	5.3	**Selected model.** Adequate statistical and visual fit, including in low exposure range. Used local fit of data. Deviance as low as best fitting models, AIC within 2 units.
Two-piece log linear spline (knot 1600)	458.1	462.6*	0.07*	yes	linear	yes	1.9	Characteristics similar to linear spline model, but that linear model was preferred.
linear, square root exposure	459.8	461.8	0.05	no	supralinear	yes	nc	Adequate statistical fit, but poor local fit at low dose.
log linear, square root exposure	460.0	462.8	0.08	no	supralinear	yes	0.20	Similar to linear, square root exposure model, less good global fit
linear, log exposure	458.2	460.2	0.02	no	supralinear	yes	nc, high	Good overall statistical fit, but very steep in low dose region
log linear using log exposure	458.4	460.4	0.02	no	supralinear	yes	nc, high	Good overall fit, lowest AIC among models. Becomes increasingly steep as exposures decrease. Preference for models with local fit.
linear, cumulative exposure	461.2	463.2	0.13	no	linear	no	0.083	Poorer fit in low exposure range, preference given to models using local fit
log linear, cumulative exposure	462.8	464.4	0.16	no	linear	no	nc, low	Poor fit: Highest deviance and AIC amongst models, 4 units above best fitting (plateau-like) models. Poor local fit to all categorical points. Not a health protective choice.

Linear models: RR = β*function(exposure); Log linear models: RR = exp(β*function(exposure)), i.e., log(RR)= β*function(exposure)
*Following SAB advice, knots were fixed in prior analyses, and not varied in determining confidence limits. As the 1600 ppm-day knot was not maximum likelihood value, it was not penalized as if it were a MLE fit.
nc: Not calculated.
nc, high: Not calculated (technical problem). Curve shape implies low dose risk well above estimates from spline 1600 models.
nc, low: Not calculated. Curve well below linear cumulative dose model. Upper bound on low dose slope` (used for unit risk calculations) is approx. 5 x lower than for linear cumulative exposure model (IRIS Appendices)

significantly better predicts the actual cases overall and the cases at lower exposures below 1600 ppm-days (which is the focus of EPA's local fit consideration). EPA's Slide 28 (Table 1) also includes a new unsubstantiated claim that the standard log linear model is "Not a health protective choice". EPA provided no scientific basis for this claim.

NASEM Committee Member Question 2: How was the knot selected?

EPA claims that EPA SAB advised EPA IRIS to use a 2-step process by first fixing the knot and then modeling the 2-piece spline model as a second step. EPA stated this in order to justify EPA IRIS p-value of 0.07 (Table 1) instead of TCEQ's corrected p-value of 0.14 (Table 2). However, EPA did not simply select a knot. EPA (IRIS) iteratively assessed every 100 ppm-day from 100 to 15,000 from the same IRIS data set in order to select a "local maximum likelihood" (EPA IRIS-D41-42). In other words, the knot was an estimated parameter in the model, requiring an additional degree of freedom.

Table 2. TCEQ (2020, p.40) corrected p-values and AIC values for TCEQ's CPH model and EPA's 2-piece spline model

Model [a]	p-value [b]	AIC [c]
Cox proportional hazards model (log-linear model)	0.22	464.4
Linear two-piece spline model with knot at 1,600 ppm-days [d]	0.14	464.5

AIC - Akaike information criteria, NIOSH - National Institute for Occupational Safety and Health

TCEQ (2020) correctly concluded that the TCEQ (2020) model has comparable p-value and AIC values (Table 2), but that the TCEQ's model is more parsimonious (1 vs 3 parameters) and consistent with the EtO-specific biological and epidemiological evidence. Instead of subjective visual fit comparisons, TCEQ uses a well-accepted approach to check if the model can predict the number of cases. In contrast, EPA has no scientifically valid basis for the 2-spline model since EPA IRIS selection is based on incorrect visual and statistical fit.

The TCEQ assessment does not include breast cancer incidence based on TCEQ's analysis of the weight of evidence for breast cancer mortality and incidence. While we agree that the weight of evidence for breast cancer causality is weak, a major consideration not addressed by TCEQ DSD is the large number of missing breast cancer incidence cases that preclude using these data for quantitative risk assessment purposes. While breast cancer mortality and incidence should be considered qualitatively in characterizing hazard, TCEQ can improve justification for excluding breast cancer incidence for quantitative risk assessment purposes.

EPA's presentation prompts us to provide comments relevant to NASEM's statement of tasks 1 and 2. We suggest Tier 1, 2 or 3 recommendations for NASEM committee members to consider.

DETAILED COMMENTS ORGANIZED BY NASEM STATEMENT OF TASKS

The following detailed comments are organized in the order of the NASEM statement of tasks (highlighted in bordered boxes) for reviewing the Texas Commission on Environmental Quality's Ethylene Oxide Development Support Document. Tier 1, 2 or 3 recommendations are highlighted in bold black font.

1. The overall weight of the evidence for a causal relationship between ethylene oxide and breast cancer risk in humans at occupational concentrations and at environmentally relevant concentrations of ethylene oxide in ambient air.

As discussed above, the TCEQ DSD relies on the meta-analysis by Marsh et al. (2019) to support their conclusion that the overall weight of evidence does not support a causal relationship between EtO and breast cancer risk in humans at occupational concentrations and at environmentally relevant concentrations of ethylene oxide in ambient air.

TCEQ could strengthen the rationale in the DSD by adding that neither the NIOSH breast cancer incidence study (Steenland et al., 2003) nor the NIOSH mortality study (Steenland et al., 2004) report an overall excess of breast cancer. The positive NIOSH findings are not robust in that they are seen with a certain lag and exposure metric that are not evident with numerous other exposure metrics, models, or lags [8]. The breast cancer incidence findings are at most suggestive, not only due to inconsistencies in the exposure response, but also due to incomplete cancer ascertainment (i.e. missing cases) and the subsequent potential for bias.

EPA IRIS breast cancer incidence analysis relied on data from the subpopulation of the NIOSH cohort that was interviewed, which required both locating subjects and identifying those diagnosed with breast cancer. Of the 7,576 women in the NIOSH cohort, only 319 cases were identified, a significant deficit from the expected number of 367 due to under ascertainment. In addition, 32% of the overall population were not included in the interview portion of the study. If all cases were found in the overall population and interviewed, the analysis would have included 134 or more breast cancer cases in addition to the 233, which became the data incorporated in the EPA model. The percent non-response was of concern, according to Steenland et al. (2003), who stated:

> "Our data suggest that EtO is associated with breast cancer, but a causal interpretation is weakened due to some inconsistencies in exposure-response trends and possible biases due to non-response and incomplete cancer ascertainment."

[8] Note: The EPA IRIS statistical modeling builds on a backdrop of a large number (>50) of statistical models already conducted in the original published papers by Steenland et al. (2003, 2004). The IRIS assessment further expands the analyses to include different

Steenland et al. (2003) indicated that cases lost are more likely to be shorter term (i.e., lower cumulative exposure) employees. If more cases were missed among those with lower cumulative exposures (shorter term employees), then the data would be biased toward seeing a positive slope and/or elevated risk in the higher exposure groups, as reported by Steenland et al. (2003).

There has been some confusion regarding Steenland et al. (2003) statement that "breast cancer ascertainment in the sub-cohort with interviews was considered complete." In other words, all the women who were interviewed were identified as having breast cancer or not. This, however, does not account for the missing cases never identified in the overall target population and among the case nonparticipants in the interview portion. Furthermore, NIOSH does not allow access to breast cancer incidence data to allow independent analyses. Thus, the Steenland et al. (2003) should be classified as useful for qualitative but not quantitative risk assessment purposes.

Tier 1 recommendation: TCEQ should consider improving the rationale for excluding breast cancer incidence for quantitative risk assessment due to the underascertainment as described above.

Tier 3 recommendation: TCEQ can be advised to revisit breast cancer mortality after peer review and publication of the recently disclosed NIOSH mortality study update. We object to premature release of these unpublished findings in the EPA presentation as basis for supporting inclusion of breast cancer incidence and application of an extremely steep dose-response model. However, if the breast cancer mortality data is made publicly available for independent analysis, then there is an opportunity for rigorous peer review and transparency of any modeling analysis performed on these data.

2. The dose-response assessment for ethylene oxide, including:

a. the appropriateness of the selected cancer endpoint(s) and key data set for modeling, dose-response model, shape of the dose-response curve, and model fit criteria, in the context of relevant mode of action information

Selected cancer endpoints(s) and key data set for modeling

Lymphoid mortality from the NIOSH cohort without further manipulation to estimate extra risk for lymphoid incidence is an appropriate health outcome to use for cancer risk assessment. However, modeling of the male lymphoid cancers alone is more appropriate based on the conclusion of the key study by Steenland et al. (2004) that the findings were in males and not females. TCEQ (2020) found that males alone resulted in a higher cancer slope factor than males and females combined. In a systematic literature review and metaanalysis, Marsh et al. (2019) concluded that the most informative epidemiology studies, which were published in the 2000s and 2010s, do not support the conclusion that exposure to EtO is associated with an increased risk of lymphohematopoietic cancer or breast cancer. This weightof-evidence is important to consider in selecting the model because there is no epidemiological evidence that EtO is a highly potent human carcinogen.

EPA commented that the 2-slope spline modeling results for all lymphohematopoietic (LH) cancer was consistent with the lymphoid results. This, however, is not an independent supportive finding, because lymphoid cancers comprise the majority of LH cancers and, therefore, drive the LH results.

One NAS committee member asked whether Hodgkin's Lymphoma was included in the NIOSH lymphoid category, as one study reported an increase in this cancer. We agree with TCEQ to continue using lymphoid mortality (which does not include Hodgkin's Lymphoma) from the NIOSH study as the cancer endpoint. It is noteworthy that there was not even one death due to HD in the UCC study of over 2,000 ethylene oxide chemical workers followed from the 1940s.

Tier 3 recommendation: EPA's presentation to NASEM indicated that there was an update for the breast cancer mortality data. TCEQ should determine if there was also an update for the lymphoid mortality data, and if not, whether this is due to lack of findings.

Dose-response model

Tier 1 recommendations: In light of EPA's incorrect presentations, we recommend that TCEQ strengthen and clarify its correct conclusion that EPA's flawed and meaningless EPA visual comparisons Relative Risk (RR) plots or new tables based on these plots (EPA, 2024) do not support EPA's inference that TCEQ's model underestimate the observed data.

During Meeting #4, EPA emphasized apparent greater risk predicted extrapolations of increased EtO-Induced lymphoid cancer risk made by EPA's 2-piece spline model vs. the TCEQ standard Cox Proportional Hazards (CPH) model. To support and re-emphasize his claim, EPA showed Figure 1, which is identical to EPA IRIS (2016) Figure 4-3 (Figure 2), shown immediately below with the footnote warning highlighted.

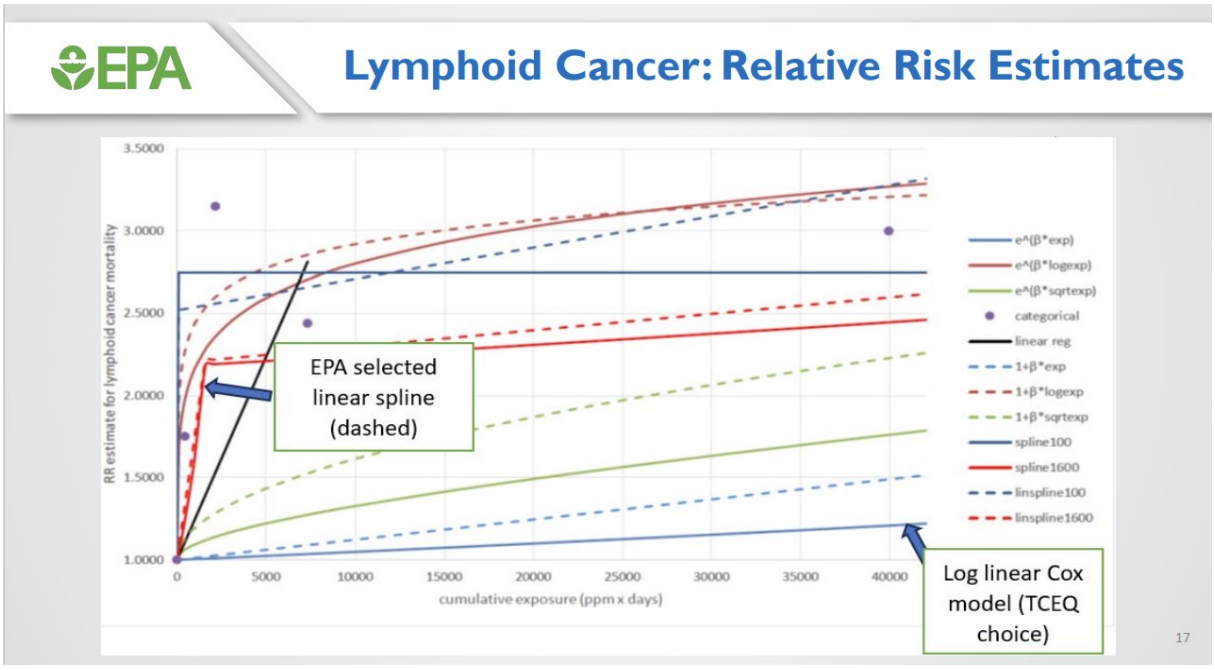
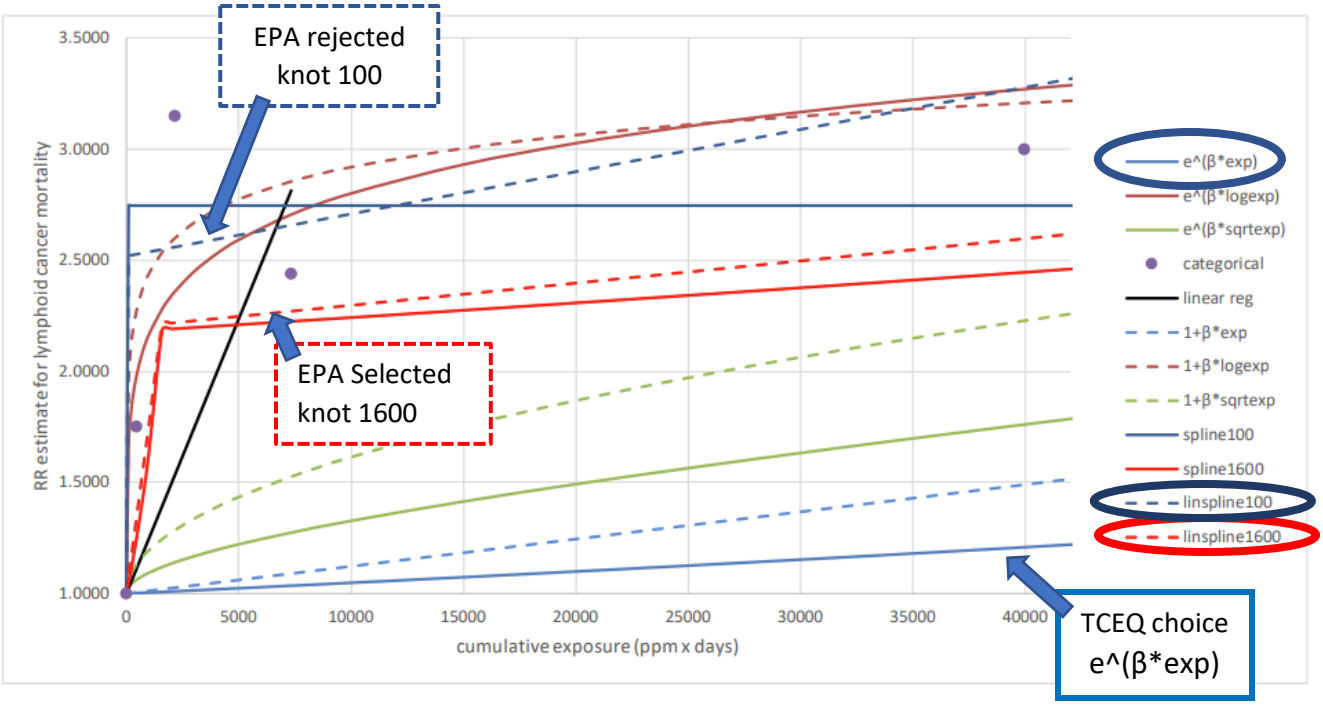


Figure 1. EPA (2024a) Slide 17 Presentation to NASEM is equivalent to EPA IRIS (2016) Figure 4-3 on p, 4-21 shown below as Figure 2.



e^(β*exp): $RR = e^{(\beta \times exposure)}$; e^(β*logexp): $RR = e^{(\beta \times \ln(exposure))}$; e^(β*sqrtexp): $RR = e^{(\beta \times sqrt(exposure))}$; categorical: $RR = e^{(\beta \times exposure)}$ 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: $RR = 1 + \beta \times exposure$; 1 + β*logexp: $RR = 1 + \beta \times \ln(exposure)$; 1 + β*sqrtexp: $RR = 1 + \beta \times sqrt(exposure)$; 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.)

Figure 2: EPA IRIS (2016) Figure 4-3

- EPA's selected 2-spline CPH model is the red dashed "linspline 1600" model with knot at 1600 ppm-days
- TCEQ's selected log-linear CPH model is the solid blue "e^(β*exp)" model at the bottom of the graph
- EPA also refers to the 2-spline CPH model with knot at 100 ppm-days which is the blue dashed "linspline 100" model that is very steep initially before bending at 100 ppm-days

Both EPA (2024a) and the EPA 2016 IRIS report strongly (but falsely) implied that the visual fit comparison shown in EPA Figure 4-3 (Figure 2 above) demonstrates that the virtually linear fit that appears at the bottom of the plot, representing the TCEQ standard CPH model fit), clearly underestimates both the plotted categorical (quartile) data points and the EPA linear spline models that incorporate a knot at a cumulative exposure of either 1,600 or 100 ppm-days. Importantly, EPA's presentation to the NASEM committee failed to mention or discuss the following clearly noted caveat provided by EPA concerning its Figure 4-3 (EPA IRIS 2016, p. 4-21, emphasis added):

"Note that, with the exception of the categorical results and the linear regression of the categorical results, the different models [shown in Figure 4-3] 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."

The caveat in the footnote of all EPA Figures of relative rate (RR) means that the **EPA Figure 4-3 presents a meaningless, invalid, and highly misleading comparison** between a virtually linear model plot (TCEQ choice) and the two EPA linear spline model fits. All models were actually fit by Steenland/EPA to the same underlying individual-level data rather than to the categorical data points plotted.

As correctly explained in detail by TCEQ (2020) at pp. 142-155, and as pointed out specifically on p. 151 of that TCEQ report, **it is TCEQ Figure 15—reprinted as Figure 3 below**, **which shows categorical as well as corresponding underlying individual-level lymphoid mortality hazard ratio data points, in contrast to the invalid and highly misleading EPA (2016) Figure 4-3—**that provides an appropriate visual comparison between EPA linear two piece spline model fits to the full (individual-level) NIOSH dataset after adjusting them for the difference in baseline risks implied by these models vs. the baseline risk implied by the standard CPH model.

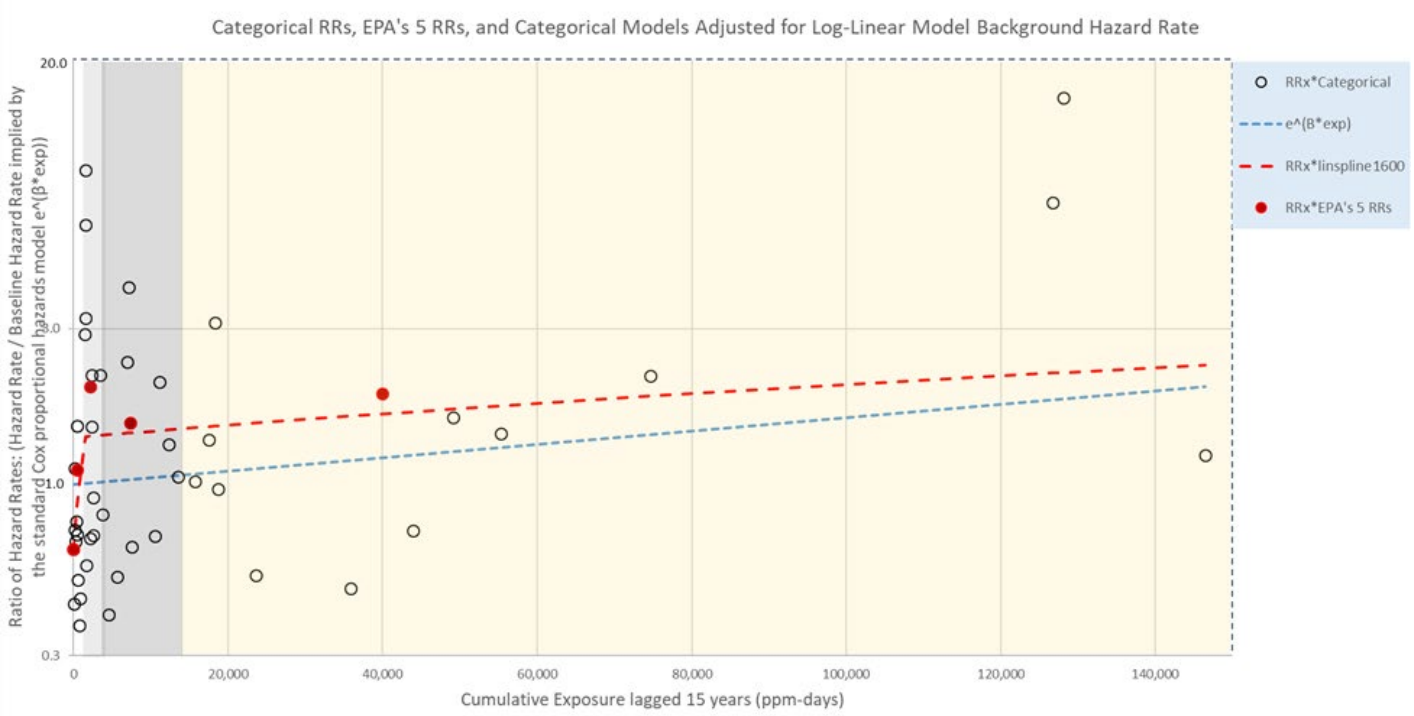


Figure 3. Figure 15 on p. 151 of TCEQ 2020, there titled "Lymphoid cancer death ratios of hazard rates estimated by the categorical RRs and the linear two-piece spline model ("knot" at 1,600 ppm-days) **after adjusting them for differences in implied background hazard rates of the standard Cox proportional hazards fitted model for 15-year lagged occupational doses ≤150,000 ppm-days (NIOSH cohort) adjusting for the difference in baseline risks between the RRs and the linear two-piece spline model".**

The y-axis in Figure 3 is no longer RR but are ratios of hazard rates after adjusting them for differences in estimated baseline hazard rate to illustrate a more valid visual comparison along the y-axis.

It is clear from the more meaningful “apples-to-apples" comparison shown in TCEQ Figure 15 that the TCEQ CPH model and the EPA linear two-piece spline model (“knot” at 1,600 ppm-days) each provide statistically comparable fits to the underlying individual-level data. In contrast, and in view of the more valid comparison shown in TCEQ Figure 15, there is no “visual fit” basis whatsoever (and certainly none by EPA Figure 4-3) to prefer the two-piece EPA spline model with a knot at 1,600 ppm-days to the (virtually linear) TCEQ standard CPH model of increased EtO-associated lymphoid cancer risk.

It is noteworthy that EPA referred the NASEM committee to EPA (2024b) response to comments which clearly illustrate EPA’s continuing misuse of these figures to incorrectly compare relative risk calculations. Instead of showing these figures, EPA (2024b) creates a new table that compares the relative risk values predicted at interval midpoints. In other words, the new table of numbers is essentially Figure 4-3 disguised as a new independent numerical analysis. This new table does not even include, let alone discuss implications of, the above-mentioned caveat that EPA had explicitly included pertaining to fundamental limitations on any valid interpretation of Figure 4-3 of its IRIS (2016) report. Thus, EPA’s continued insistence, including to NASEM during its Meeting #4, that it developed any visual-fit basis favoring its 2-piece spline model over the TCEQ CPH model of increased EtO-related lymphoid cancer risk is inexplicable and has no objective, scientific basis.

In terms of future recommendations, the burden is on EPA to provide a published reference where comparisons of Cox proportional hazard models fit to individual data are done using visual fit comparisons with categorical estimates.

However, as a Tier 2 recommendation, we strongly urge TCEQ to address EPA’s (2024b) new response to comments by explaining that EPA’s new table of values is essentially Figure 4-3 in the guise of an independent quantitative analysis. This also applies to EPA’s Reference Slide 30 in the EPA (2024a) June 27, 2024 presentation to the NASEM Committee. TCEQ could provide a reference supporting the fact that the y-intercept is estimated based on the type of model. EPA appears to assume incorrectly that the y-intercept for continuous models represents the lagged out group just like the categorical estimates. As a result, EPA ignores the footnote warning in the IRIS assessment.

Shape of the dose-response curve

EPA never responded to comments related to reconciliation of choice of a low dose steep model, i.e., indicative of a potent carcinogen with the findings of the NIOSH studies that do not imply such a pattern. The original Steenland et al. (2004) mortality study reports no effect in females and an increased risk of lymphoid cancers in males only at the highest exposure level. The studies do not show a low dose increase. Furthermore, the Agency never addressed the absence of increases of lymphoid cancers in both external and internal analyses of the UCC study of males. EPAs choice for a continuous model hinge inappropriately on 5 categorical odds ratios from the NIOSH study for males and females combined.

The EPA (2024a) Presentation Slide 9 cited Mikoczy et al. (2011) study as supportive of low dose steep risk minimizing serious concerns that have been raised regarding the unusual breast cancer deficit in the internal comparison referent group. The Agency also dismisses relevant data showing absence of HWE for breast cancer. A referent group whose breast cancer rates are 50% of general population raises issues that can't be explained by the HWE. HWE is primarily related to non-cancer causes and declines with length of follow up. An internal referent group with an unusual deficit of the disease of interest creates an artifact of an excess as is illustrated in this study.

Model fit criteria

Tier 1 Recommendation: The following edits to TCEQ (2020) could strengthen and clarify its correct conclusions concerning flawed EPA statistical analysis of AIC and P-values for Goodness of Fit by EPA's 2-Piece Spline Model vs. by the TCEQ standard Cox Proportional Hazards (CPH) Model

TCEQ (2020) at pp. 138-139 clearly explains how the EPA IRIS (2016) EtO risk assessment ignores standard statistical practices. **TCEQ could consider the following edits in bold brackets.**:

- The "knot" was an iteratively fit model parameter and was not simply preselected (p. 4-52 of USEPA, 2016); and

- The knot values, being statistically estimated/optimized based on the NIOSH data, clearly do not conform to the USEPA SAB's notion of potentially fixing some model parameters *not estimated from the data* in the interest of parsimony (see p. 12 of EPA SAB, 2015).

For the spline models there were **[not 2, as assumed in EPA IRIS (2016), but rather a total of k = [9]]** 3 parameters … estimated by USEPA: (1) the "knot" value; (2) the slope above the knot; and (3) the slope below the knot …. However, USEPA (2016) did not account for statistically estimating the optimized knot value. Thus, it appears the degrees of freedom (df) were inappropriately reduced for the spline models (df=k, the number of additional parameters estimated for this model over the model with zero-slope with cumulative exposure). This was not inconsequential. Among other consequences, this **[erroneously[9]]**:

• Decreased the p-value for adequate statistical fit, incorrectly implying that the linear two-piece spline model with a knot at 1,600 ppm-days for lymphoid cancer fit the data statistically better than other models in Table 4-6 of USEPA (2016); and

• Decreased the AIC for the spline models, which did not allow for an appropriate comparison of model fit.

…

… a p-value of 0.14 is the correct p-value for the likelihood ratio test (not 0.07 as in Table 4-6 o USEPA 2016) when appropriately using k=3 for the linear two-piece spline model with a knot at 1,600 ppm-days for lymphoid cancer. Thus, the correct p-values indicate that the likelihood of the linear two-piece spline models with a knot at 1,600 ppm-days is not different from the likelihood of the null model at the 5% significance level (i.e., the fitted two-piece spline models do not explain the variability in the data statistically significantly better than the null model). The log-linear (standard Cox regression) model has a similar p-value (0.22) and also does not explain the variability in the data statistically significantly better than the null model (Table 5). Thus, the appropriate calculation of p-values puts TCEQ's preferred model for lymphoid cancer (i.e., the Cox model) and the model used by USEPA (2016) (i.e., the linear two-piece spline model) on equal ground in this regard, although TCEQ's preferred model is more parsimonious **(i.e., uses 1 parameter compared to EPA's 3 parameters[9])**.

Pages 140-141 of the TCEQ 2020 report further clarify that

This knot was determined so that the likelihood of the spline model was maximized. That is, the knot [estimated by EPA at 1,600 ppm-days] is another parameter that was searched for outside [EPA-implemented calculations using] SAS [statistical software]. 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 that SAS reported

[9] Bracketed text denotes suggested additions TCEQ can add for clarity

> does not reflect the fact that the knot was also estimated. The Chi-Square that accounts for the fact that the knot was estimated outside SAS should then be 5.412, with k (the degrees of freedom) being three. This corrected calculation results in a p-value of 0.1440. That is, the p-value (0.14 in Table 5 [**on p. 40 of the TCEQ 2020 report, in contrast to the erroneous p-value of 0.07 for the EPA linear spline model listed in Table 4-6 on p. 4-19 – 4-20 of EPA IRIS 2016**[10]]) indicates that the likelihood of the linear two-piece spline model with a knot at 1,600 ppm-days (preferred by USEPA 2016) is not different from the likelihood of the null model at the 5% significance level. In short, there is no evidence indicating that the fitted linear two-piece spline model explains the variability in the data any better than the null model.

A related discussion on p. 41 of the TCEQ 2020 report likewise points out that

> Table 5 shows that the linear two-piece spline model with a "knot" at 1,600 ppm-days used by USEPA (2016) does not fit the data statistically significantly better than the null model (zero slope) at the 5% significance level (i.e., the linear two-piece spline model does not explain the variability in the data statistically significantly better than the null model). Likewise, the standard Cox regression model preferred under TCEQ (2015) does not fit the data statistically significantly better than the null model. Additionally, the AIC values for the Cox and the linear spline models are similar. Thus, based on standard statistical model fit criteria (i.e., p-values and AIC values), neither model provides a statistically superior fit to the modeled individual lymphoid cancer mortality data.

Tier 1 recommendation:

TCEQ's Table 5 on p. 40 should include the number of parameters in the model (1 for the log-linear model and 3 for the 2-piece spline model).

The TCEQ discussion could be augmented to point out that, according to criteria stated in footnote b of Table 4-6 on p. 4-20 of the EPA IRIS (2016) report, the corrected p-value of 0.14 for the linear two-piece spline model with a knot at 1,600 ppm-days (preferred by EPA, 2016) indicates an "inadequate" (i.e., neither a "good" [$p < 0.05$] nor an "adequate [$0.05 < p < 0.10$]) statistical fit to the modeled lymphoid cancer mortality data. So neither the EPA linear two-piece spline model with a knot at 1,600 ppm-days nor the TCEQ CPH model provide evidence that the modeled lymphoid cancer mortality data exhibit an adequately significant positive trend using EPA fit-adequacy criteria.

Typically such a cancer endpoint would be eliminated from consideration for cancer risk extrapolation based on individual-level epidemiological data. For example EPA (2012)

[10] Bracketed text denotes suggested additions TCEQ can add for clarity

Benchmark Dose (BMD) Technical Guidance (EPA/100/R-12/001) at pages ix and 15 states that "a minimum dataset" for BMD dose-response modeling requires that "There should be at least a statistically or biologically significant dose-related trend in the selected endpoint." The rationale for this requirement is to ensure that risk extrapolation models should not be fit to data that are consistent with statistical noise rather than a meaningful elevation in risk.

Tier 1 recommendation: The TCEQ text on pages 138-139 and 140-141 should be corrected to note that the knot contained in the EPA linear spline model with a knot at 1,600 ppm-days in fact *does not* represent an iteratively optimized maximum likelihood estimate (MLE) of the knot incorporated into this EPA-preferred spline model.

The reason for our Tier 1 recommendation is that EPA's presentation to NASEM pointed to this fact to dispute the TCEQ claim on pages 140-141 of its report that EPA IRIS (2016) reported erroneous associated values of model-specific p-value for fit and AIC.

The EPA IRIS (2016, Appendix D, at page D-42) report makes clear that the EPA-preferred two-piece linear spline log RR contains a *local maximum likelihood*, not a *true MLE*, knot at 1,600 ppm-days (emphasis and clarifying text added in bold italics):

> Model results for the categorical and two-piece linear log RR models are shown in Tables D-29 and D-30. Tables D-31 and D-32 give the results for the log-transform model and linear log RR models; the latter does not fit the data well. Table D-33 shows the model results for the two-piece log-linear spine model ***with the knot at the local maximum likelihood of 1,600 ppm-days.***
>
> Figure D-16 shows the graphical results for the categorical, (log-) linear, two-piece (log-) linear, and log-transform log RR models. There is a very steep increase in risk at very low exposures. The ***[true MLE]*** knot for the two-piece log-linear curve is a low 100 ppm-days. The steep slope at low exposures may be unrealistic as a basis for risk assessment, dependent as it is on relatively sparse data in the low-exposure region. Table D-34 lists the cumulative exposures with a 15-year lag for all the lymphoid cancer cases (e.g., there are no cases below the knot of 100 ppm-days).

Note that the Y-axis plots –2×(the log likelihood) of the data, and that the maximum likelihood (and thus also the maximum log likelihood) of the data by definition occurs at the minimum value of –2×(the log likelihood) of the data, which occurs at a cumulative exposure value of 100 ppm-days. This is shown in Figure 4 below, which is adapted from EPA IRIS (2016, Appendix D) Figure D-15 that is associated with the EPA (2016, Appendix D) discussion quoted just above. Figure 4 also shows that the 1,600 ppm-day knot occurs at a local maximum likelihood (plotted as the 2nd minimum negative log likelihood value as that negative-likelihood function increases along the Y-axis from left to right along knot values plotted on the X-axis).

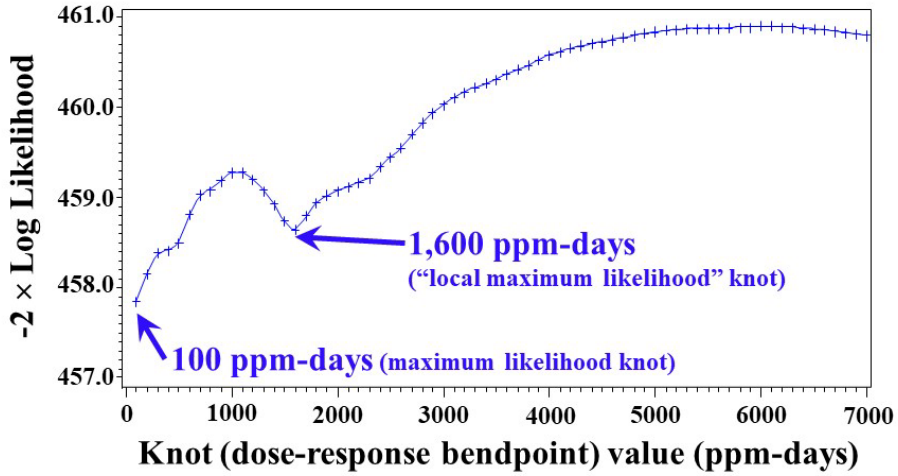


Figure 4. Adapted from: EPA IRIS (2016, Appendix D) Figure D-15 (p. D-42): Likelihood vs. knot value for 2-piece log-linear model, lymphoid cancer mortality.

Tier 1 recommendation: TCEQ could add the following discussion and consider adding the EPA figure (Figure 4): Thus, while EPA is correct that the 1,600 ppm-day knot preferred by EPA (as being more biologically realistic) does not represent an estimated maximum likelihood knot value, it is clear that numerical estimation procedure used to identify the EPA-preferred knot as the next-highest local minimum along the plotted likelihood curve resulted in a model knot-parameter value estimated *directly* and *quantitatively* from the modeled data, and thus reflects an estimation procedure that *requires* a degrees-of-freedom adjustment associated with the estimated model as otherwise correctly explained by TCEQ. There is no valid statistical justification to ignore this basic statistical-procedure requirement. Thus, EPA is incorrect to claim that EPA SAB instructed EPA IRIS (2016) to perform the analysis in two steps by fixing the knot. EPA did not fix the knot; it estimated the knot *directly* and *quantitatively* from the modeled data

Tier 2 recommendation: TCEQ could add that this basic principle of accounting for all modeled parameters is clearly stated in the National Research Council report entitled "Models in Environmental Regulatory Decision Making", that the strategy to pick the "best model" for regulatory decision making should be "subject to a penalty function reflecting the number of model parameters, thus effectively forcing a trade-off between improving model fit by adding addition[al estimated] model parameters versus having a parsimonious description" (NRC, 2007, pp. 174).

Mode of action information

TCEQ includes a good discussion of the putative genotoxic mode of action and other toxicology and pharmacokinetic data specific to ethylene oxide that informs selection of the standard CPH model. However, it may be advisable for TCEQ to further strengthen this evaluation in light of the EPA (2024a) presentation at the June 27, 2024 NASEM meeting.

At that meeting, EPA referenced EPA's (2022) public comments which EPA claimed showed that there is no clear biological evidence supporting either the EPA's or TCEQ's model. EPA (2022) conducted a highly subjective visual inspection of genotoxicity and cancer bioassay data to support their claim that the biological evidence cannot be used to inform biological plausibility. The EPA (2022) evaluation involved (a) plotting the data as point estimates without error bars, (b) drawing a straight line between the response levels for the lowest and highest dose levels, and (c) declaring the exposure response to be supralinear or sublinear depending on whether the responses for the mid-dose levels visually appeared to be above or below the line. This visual inspection did not involve any consideration of statistical significance or evaluation of which data set and dose regimen is most relevant and useful to inform the shape of an exposure-response curve based on epidemiology data. Based on this analysis, EPA argued to the NASEM committee that the exposure response information obtained from animal genotoxicity and carcinogenicity studies would not allow selection of exposure response models for human risk assessment for EtO.

This position seems to be at odds with the agency's 2005 carcinogen risk assessment guidelines stating that "[i]f dose-response analysis of nontumor key events is more informative about the carcinogenic process for an agent, it can be used in lieu of, or in conjunction with, tumor incidence analysis for the overall dose-response assessment." Except for studies investigating sister chromatid exchanges (SCE), which are no longer considered as reflective of investigating genetic damage (Wilson and Thompson, 2007), the exposure response information for other genotoxicity endpoints from animal studies should be considered in the selection of the model for EtO risk assessment. The putative genotoxic mode of action and toxicokinetic data should be a major consideration when selecting a model for risk assessment. While there is clear evidence that EtO is genotoxic and an animal carcinogen, a genotoxic MoA should be

considered as a default assumption in the absence of a convincing alternate MoA that does not involve genotoxicity as the initial key event. For a direct acting alkylating agent such as EtO, the default exposure response for the induction of mutations is linear. This is the worst-case scenario since at low doses closer to the origin, one should expect cellular protective mechanisms (e.g., detoxification and DNA repair) to offer protection, resulting in a shallower slope in this region when compared to higher doses. Based on a presumed genotoxic MoA, both TCEQ and EPA/OEEHA estimate cancer risk based on a linear extrapolation from the point of departure (POD) to the origin but apply very different statistical models to the same epidemiological study to derive the POD, i.e., Cox proportional hazards (CPH) model by TCEQ vs. the two-piece spline model by the EPA/OEEHA. In the 2-two-piece spline model, the initial slope rises rapidly at lower exposure levels and then rises more gradually for higher exposures.

This type of exposure response is not consistent with the biology of how EtO is hypothesized to work as a direct acting genotoxicant as evidenced from studies conducted in rodents. In their response to public comments, the EPA (2022) states that "EPA does not agree that observed exposure response pattern at high dose in rodent genotoxicity or carcinogenicity studies will generally be predictive of exposure response patterns in humans." While it is true that the exposure response observed at high doses in rodent studies often overestimate the hazard potential at low, environmentally relevant exposures, they seldom underestimate the hazard at the lower exposure levels for a vast majority of chemicals and most certainly for direct acting (i.e., those that do not require metabolic activation) DNA-reactive agents.

Secondly, even in those so-called high dose studies, there are often lower dose levels with no detectable effects supporting the argument that the slope at the lower end of the exposure response cannot be steeper than at the higher doses. The exposure response curves for EtO-induced gene mutations in the bone marrow (Recio et al., 2004) and lung (Manjanatha et al., 2017) tissues of transgenic Big Blue mice illustrate the above point.

Tier 1 Recommendation: The TCEQ DSD presented the Recio et al. (2004) data. TCEQ can add the Manjanatha et al. (2015) data for additional emphasis with added explanation that these two tissues are selected to illustrate the exposure response because they represent targets for EtO-induced tumors. In both cases, there is no evidence for a steeper initial slope at the low end. In fact, the mutant frequency at 25 and 50 ppm EtO was not statistically different from the concurrent negative control group in Recio et al. study whereas only the 200 ppm, but not 100 ppm, was significantly elevated from the control in the Manjanatha et al. study. The initial shallow exposure response pattern is both of these studies is more consistent with the CPH model than the 2-piece spline.

2b. the model accuracy and validation analyses for lymphoid cancer and the conclusions based on those analyses, as well as the additional analyses that assumed a healthy worker effect for lymphoid cancer mortality

TCEQ's prediction analysis provides an important reality check that is based on well-accepted methods to check how well the model is able to predict the number of cases. It is a superior, objective basis for selecting their model, compared to EPA's misleading visual presentation of data and incorrect statistics (described above). TCEQ essentially checks the two models by using the models and national mortality rates to estimate the number of cancer deaths that each model would predict in the NIOSH study.

TCEQ (2020) showed that the TCEQ model accurately predicts the number of mortalities that were observed in the NIOSH study. In contrast, the EPA model significantly overestimates the expected cancer cases in the NIOSH cohort. The standard log-linear Cox Proportional Hazards (CPH) model used by TCEQ (2020) estimated 52.4 lymphoid mortalities compared to 53 observed mortalities. By contrast, the EPA's 2-piece linear spline model with a knot estimates 91.7 mortalities. When cast in terms of excess mortalities, the overprediction by EPA is even more stark.

One possible reason that EPA has offered to explain its higher predictions of excess cases compared to background rate is that there is a HWE in the cohort; and, therefore, the expected cases without ethylene oxide exposure are fewer than would be predicted by the life tables. With lower expected cases from background, a greater portion of the observed cases are assumed to be due to EtO exposure.

Steenland et al. (2004) concluded that there is no HWE in the NIOSH cohort based on "all-causes" mortality (i.e. cancer and all other causes) where the observed deaths of 2852 were 90% of the values expected from the life tables (Steenland et al. 2004). For "all cancers" mortality the 860 observed deaths were 98% of the values expected from the life tables. These results are consistent with the generally understood concept that the HWE "is known to vary with type of disease, being smaller for cancer than for other major diseases, and it tends to disappear with time since recruitment into the workforce." (IARC, 1999). Thus, the observed data are not consistent with a HWE for cancer, which is consistent with generally accepted epidemiologic principles. However, EPA (2022a, p. 90) points to the SMR of 0.72 in Steenland et al. (2004) Table 3 in support of an even larger potential HWE (i.e., 28%) in the NIOSH cohort for LH cancer deaths. This data unreliable because it is based on a very small sample size, 9 cases in males and 2 in females. It is also inconsistent with Steenland et al (2004) conclusions that there is no HWE for cancer in this cohort.

Thus, TCEQ assumed a reasonably high estimate of 15-16% for a HWE in a sensitivity analysis to illustrate the EPA model still overestimates the excess cases even when assuming a HWE as a

surrogate for potential differences between the general population and the NIOSH worker cohort. The observed excess cancer mortalities are 10.4. The EPA model predicts 35.0 and 76.4 excess cancer mortalities, respectively, for the central tendency and upper-bound. These values are 340% and 740% above the observed excess cases, respectively. Thus, the EPA model still substantially overpredicts the excess cancers even assuming a reasonably large HWE.

Tier 1 Recommendation: EPA's new reference slide 29 criticizes the TCEQ "reality-check" by questioning TCEQ's method for estimating the confidence intervals (CIs). EPA applies a double standard to reject TCEQ's analysis based on method of calculating CIs when EPA IRIS relied on inappropriate visual fit comparisons with categorical estimates without any consideration of CI's. TCEQ should emphasize that TCEQ's model predicted 14.4 cancers in Quintile 2 vs actual 11. Quintile 2 (up to 1550 ppm-days) best reflects EPA's definition of "local" fit below the 1600 ppm-day knot for EPA's model. Clearly, TCEQ's model does not underestimate the cases at the lower exposures.

Tier 2 Recommendation: To address EPA's criticism regarding method of calculating CI's, TCEQ could consider applying a second approach for estimating the 95% CI using the Exact Poisson method as shown in Table 6.

Table 6 Quintile-Specific NIOSH Cohort Lymphoid Cancer Mortalities Predicted by Cox and Linear Two-Piece Spline Models (Confidence Intervals Based on Exact Poisson Distributions.

Model	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Lymphoid Cancer Deaths Observed in NIOSH Cohort	*11*	*11*	*11*	*11*
Standard Cox model – 15-yr lag (MLE)	14.4 (8.1, 28.9) [11] (7.7, 23.5) [12]	8.0 (4.5, 16.1) (3.5, 15.8)	9.4 (5.2, 18.8) (4.1, 17.1)	9.1 (5.1, 18.3) (4.1, 17.1)
Linear two-piece spline with knot @ 1,600 ppm-days – 15-yr lag (MLE)	**20.9** **(11.7, 42.0)** (12.2, 30.9)	17.6 (9.8, 35.2) (9.9, 27.2)	**20.8** **(11.6, 41.7)** **(12.2, 30.9)**	**20.9** **(11.7, 41.9)** **(12.2, 30.9)**

[11] TCEQ (2020a) used the inverse of the confidence intervals of the SMRs.

[12] For comparison, we calculated the CI based on Exact Poisson distribution(Ahlbom, 1993)

2c. any implications of the endogenous production of ethylene oxide, including from the perspective of biological significance, for risk-based air concentrations

Both the CPH model and IRIS model estimate extra risk. This fact should not be used as a basis to ignore reality checks based on valid estimates of endogenous levels. While EPA's potency estimate technically only applies to exposures above endogenous levels, it is implausible that a chemical would be a potent carcinogen at levels at and substantially below levels that the body produces through natural processes.

The Kirman analysis (Kirman and Hayes, 2017; Kirman et al., 2021) of hemoglobin N-(2-hydroxyethyl)-valine (HbHEV) adducts provides an important reality check on the IURs modeled from epidemiological data. TCEQ (2020; Appendix 5) only includes Kirman and Hayes, 2017. EPA's presentation to NASEM dismisses the Kirman analysis as unvalidated. This criticism of Kirman and Hays (2017) is outdated and misleading. New datasets including NHANES HEV measurement made in >4500 individuals. Adjustments for ambient exposures based on EPA measurement has been addressed in Kirman et al. (2021) in which the exogenous pathway for EtO in air was explicitly included in the estimates of endogenous exposures to EtO.

A Tier 1 recommendation is that TCEQ can further strengthen this section to include new analysis by Kirman et al. (2021) that further informs on the shape of the exposure-response curve (TCEQ, 2020; page 39), and indicates a linear relationship between external EtO exposure and internal EtO dose.

Specifically, TCEQ can highlight four pieces of evidence strongly supported the Kirman analysis that showed a linear relationship between HEV and EtO exposure:

(1) PK model predictions of DFG 1994;

(2) PBPK model predictions of Filser and Klein 2018;

(3) empirical data collected from exposed workers across a broad range of concentrations, not just high concentrations (depicted in Fig. 12A of Filser and Klein 2018); and

(4) empirical data from NHANES smoker HbHEV data that support a linear relationship between the exposure measure of cigarettes per day (CPD) and measured HbHEV, a relationship that has been reported in multiple studies of smokers (Tornqvist et al. 1986; Bailey et al. 1988; Bader et al. 1995; Boogaard et al. 1999; Wu et al. 2004) (Figures 5,6; Note – log-log scale used in Figure 6 permit visualization of the NHANES data, all lines depicted in this figure are linear with CPD).

Importantly, after converting between CPD to equivalent air exposures, the slopes for HbHEV observed in workers and in smokers are consistent with one another. Collectively, these data provide strong validation of a linear relationship between EtO exposure and HbHEV.

Figure 5. Three Lines of Evidence Supporting a Linear Relationship Between EtO Exposure and HEV

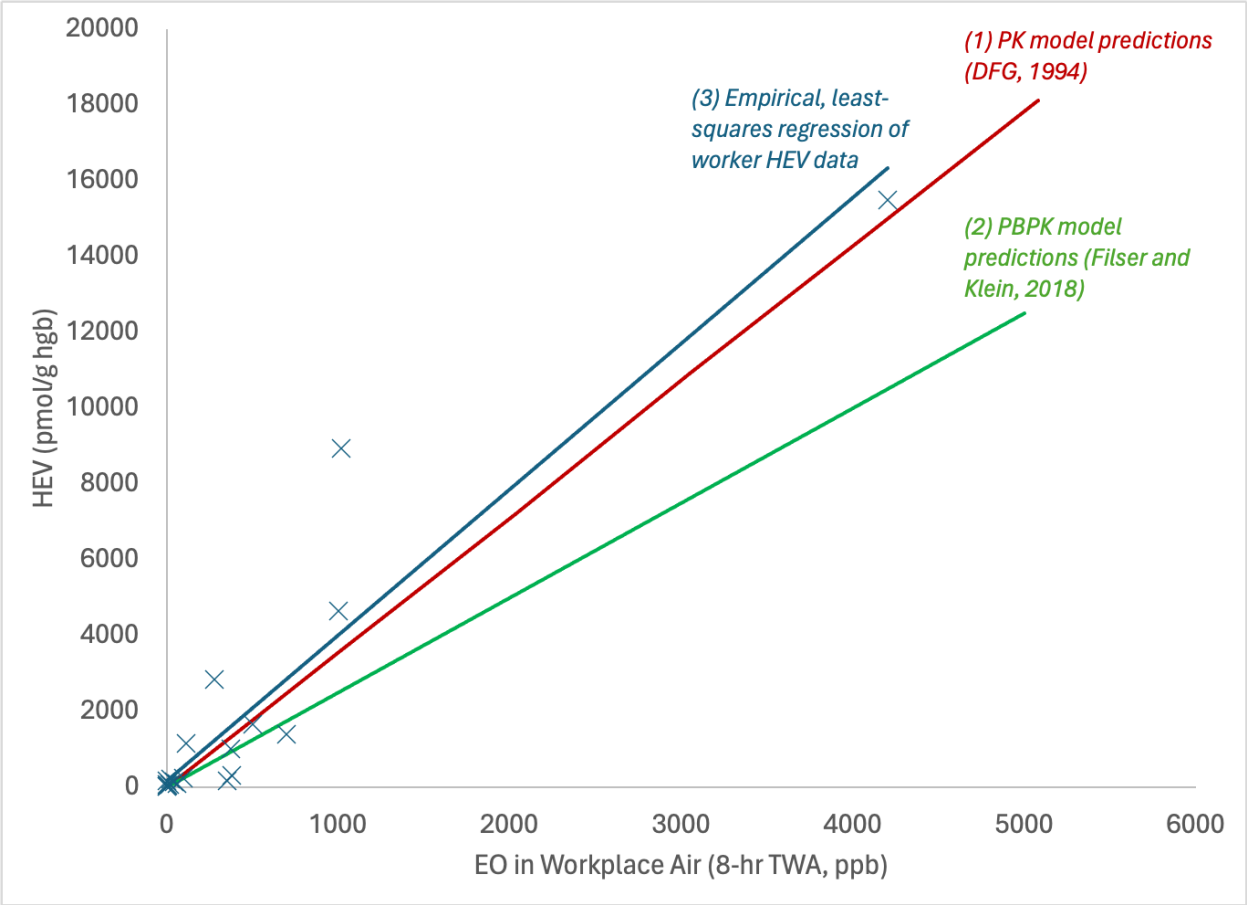
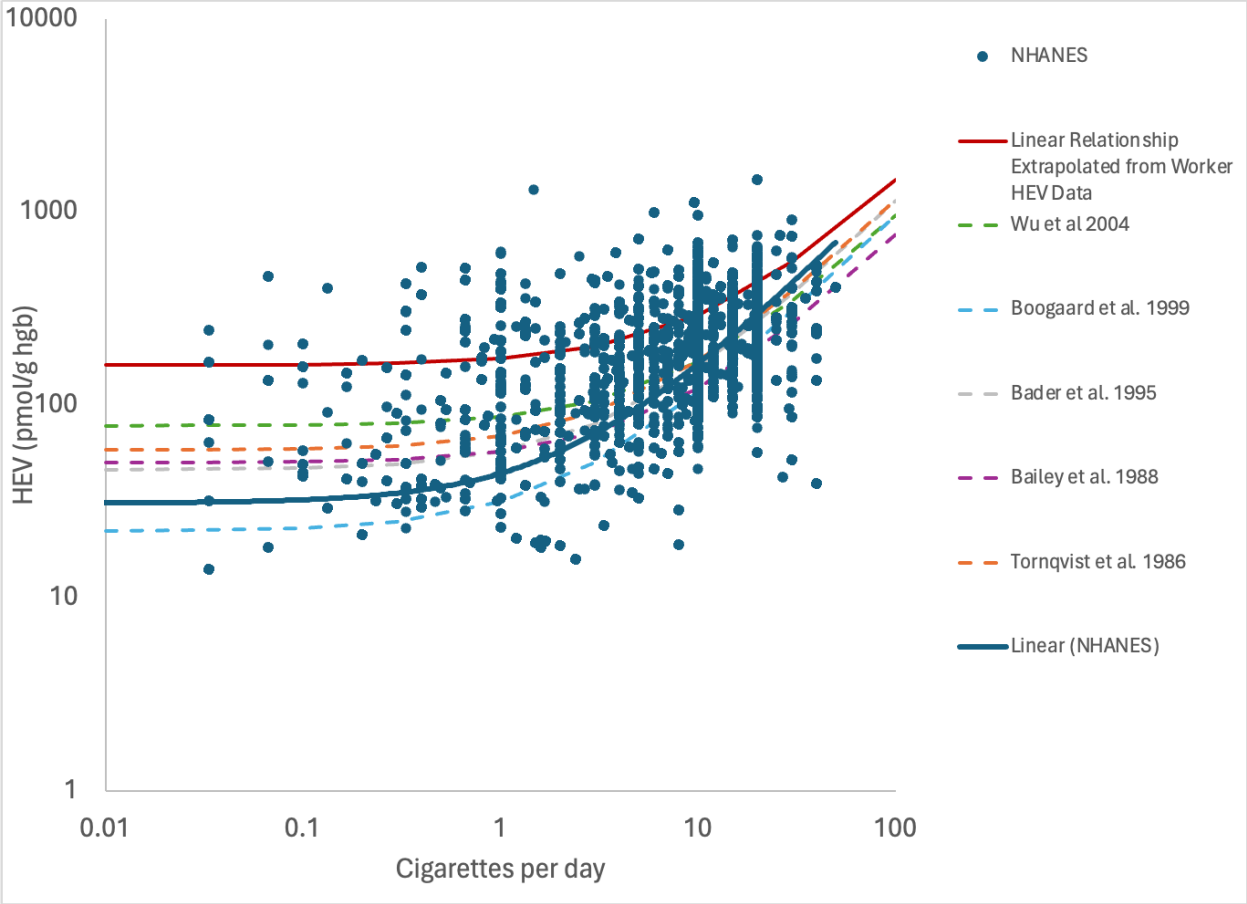


Figure 6. Fourth Line of Evidence Supporting a Linear Relationship Between EtO Exposure and HEV



EPA's failure to embrace linear toxicokinetics for EtO and HEV at low and endogenous exposure levels reflects an ironic departure from standard practice, and their attempt to hold out due to "uncertainty" arising from some as-yet-to-be-identified source of nonlinear toxicokinetics is part wishful thinking (in the hopes of avoiding the inconvenient implications of our assessment of exogenous vs endogenous exposures) and part toxicokinetic denialism.

EPA's presentation to NASEM (Slide 21) selectively showed their response to comments (RTCs) for the Sterilizer RTR (2024c) but omitted an important independent analysis attached to ACC's comments on that rule and expanded upon in the ACC's 2023 HON comments that supported Kirman's HbHEV exposure model. Indeed in the Sterilizer's RTC EPA simply asserted that no such data measuring EO in tobacco The comment (ACC, 2023; page 40) fulfilled EPA (2022)'s own recommendation for a "forward determination" as a validation of the Kirman's exposure model. This independent validation analysis relied on direct quantitation of EtO in cigarette smoke and estimates of daily smoker EtO inhalation with associated dose-dependent correlations to NHANES reported smoking-intensity HbHEV adducts. As shown in those comments there was close agreement between: (1) the original Kirman analysis; (2) NHANES

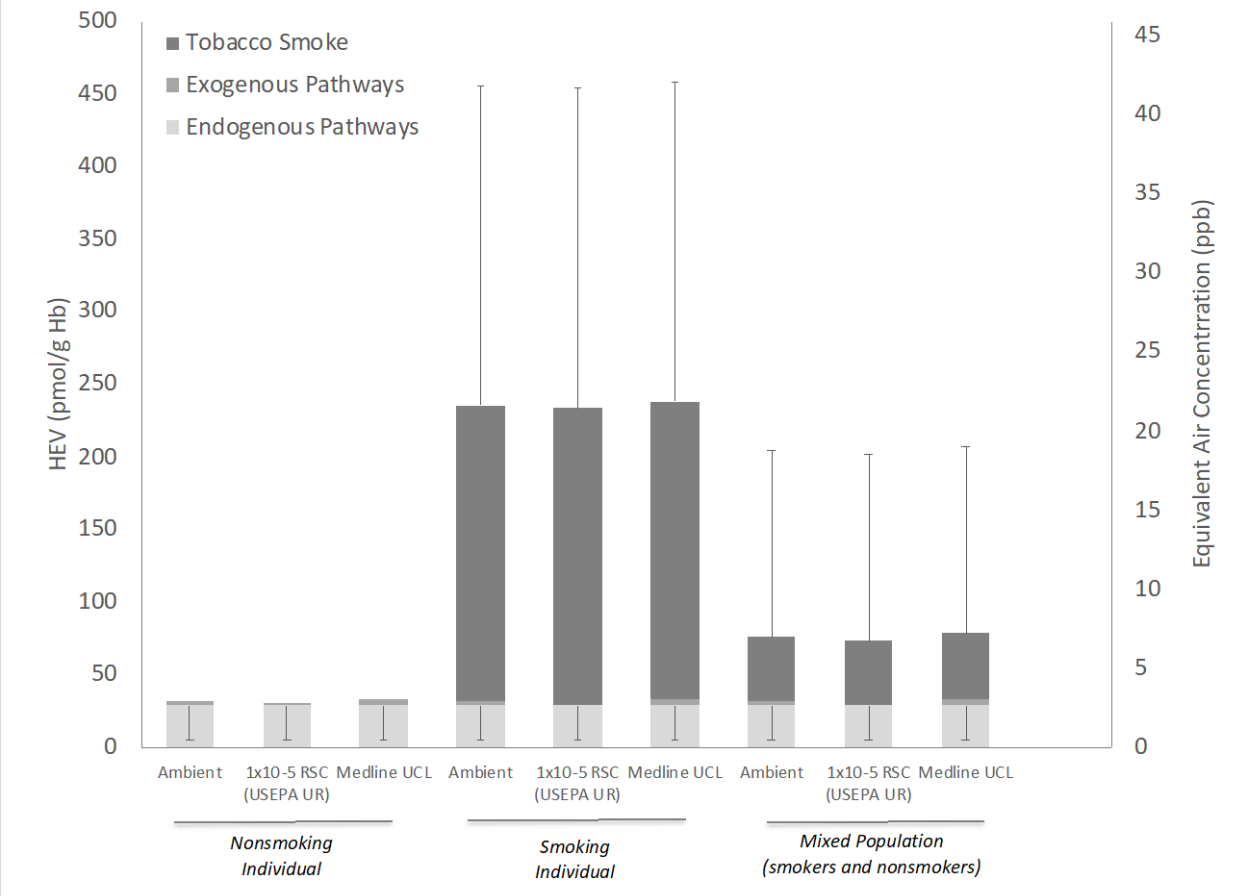
smoker HbHEV data (measuring smokers and non-smokers); and (3) the forward extrapolation from measurements of EO in cigarette smoke. Thus, validating the relationship between EO exposure and HbHEV levels.

EPA (2024a), instead, described on Slide 22 of a poster presentation made on June 5, 2024, that challenged the Kirman analysis. The poster presentation is unavailable for a review but the EPA's reliance on mean ethylene breath concentration~0.5 ppb to establish a lack of correlation to HbHEV strikingly contrasts with orders-of-magnitude higher ethylene exhalation values reported in the literature, suggesting uncertainty in the EPA ethylene-exhalation based validation challenge of the Kirman HbHEV-based EtO exposure model. As discussed in Kirman et al. (2021), ethylene production and resulting ethylene biomarkers are highly variable in humans depending upon many gastrointestinal lumen and systemic factors, as well as inter-study variation due to differences in analytical methods. EPA's focus on the study of Paardekooper et al. (2017) is puzzling for several reasons. First, ethylene present in exhaled breath is a highly variable biomarker of ethylene exposure, and highly dependent upon the study design and analytical methods. Second, there appear to be better studies than Paardekooper et al. (2017) in the published literature that characterize ethylene in exhaled breath. For example, closed chamber studies of Filser et al. (1992) and Denk (1990) that measure the accumulation of ethylene produced by unexposed human volunteers over hours (as compared to collection of 2 exhaled breaths in the Paardekooper study) would provide more robust (and higher) estimates of ethylene in exhaled breath. There are also other studies in the published literature that support higher concentrations of ethylene in exhaled breath that EPA has not considered here. Lastly, low concentrations of ethylene in exhaled breath are not incompatible with significant endogenous exposures to EtO since one of the reasons that exhaled ethylene may be lower in some individuals is that their metabolic clearance of ethylene (i.e., ethylene oxide production) is higher than other individuals. As such, EPA's selective use of ethylene biomarker data do not detract from the HbHEV biomarker data or from our estimates of endogenous EtO exposure that rely on them.

Because EPA's criticisms do not detract from the high confidence in our endogenous exposure estimates for EtO, our conclusions from our 2021 paper remain intact (Kirman et al., 2021). Specifically, potential risks from ethylene oxide are expected to be related to total exposure (i.e., from exogenous and endogenous pathways combined). When the exposures via the exogenous pathway are viewed within the context of total exposure to ethylene oxide in nonsmoking individuals, smoking individuals, and to a mixed population of smokers and nonsmokers, it can be concluded that the exogenous exposure pathway in the vicinity of these facilities is not an important contributor. Comparison of the total exposures for different exogenous exposures (ambient levels, $1x10^{-5}$ RSCs, site-related concentrations based on Medline IL site) in Figure 7 can be used to support a conclusion that site-related exposures do

not contribute to a "substantial difference" from background total exposure that would warrant additional attention/action. That is, the column heights for each of the three exogenous exposure options are essentially the same, particularly when the variation in the endogenous and tobacco smoke pathways (error bars = SD in Figure 7) are taken into consideration. Since increased prevalence of cancer was not observed with the elevated exposures to ethylene oxide from smoking (dark grey regions in Figure 7) (Jain, 2020), and since the exogenous pathway (medium grey regions in Figure 7) is not an important contributor to total exposure in most cases, managing the potential risks from the exogenous pathway by investing resources to reduce ambient or site-related air concentrations of ethylene oxide to $1x10^{-5}$ RSC levels are not expected to result in a meaningful change to public health for nonsmoking individuals, smoking individuals, or a mixed population of smokers and nonsmokers.

Figure 7. Comparison of Ambient Background, 1×10^{-5} Risk-Specific Concentration, and Site-Related Air Concentrations to Total Exposure to Ethylene Oxide. UR = Unit Risk, UCL = Upper Confidence Limit.



REFERENCES

Bailey E, Farmer PB, Shuker DE. 1987. Estimation of exposure to alkylating carcinogens by the GC-MS determination of adducts to hemoglobin and nucleic acid bases in urine. Arch Toxicol 60: 187–91. PMID: 3619640. doi:10.1007/BF00296978.

Bader M, Lewalter J, Angerer J. 1995. Analysis of N-alkylated amino acids in human hemoglobin: Evidence for elevated N-methylvaline levels in smokers. Int Arch Occup Environ Health 67: 237–42. PMID: 7591184. doi:10.1007/BF00409405.

Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, Vickers C, Willcocks D, Farland W. 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit Rev Toxicol 36(10): 781–792. https://doi.org/10.1080/10408440600977677

Boogaard P J, Rocchi PS, van Sittert NJ. 1999. Biomonitoring of exposure to ethylene oxide and propylene oxide by determination of hemoglobin adducts: Correlations between airborne exposure and adduct levels. Int Arch Occup Environ Health 72: 142–50. PMID: 10392561. doi:10.1007/s004200050353.

Denk B. 1990. Abscha¨tzung des kanzerogenen Risikos von Ethylenoxid fu¨r den Menschen durch Speziesextrapolation von der Ratte unter Bru¨cksichtigung der Pharmakokinetik. GSF-Bericht 20/90, Munchen.

DFG. 1994. Deutsche Forschungsgemeinschaft. Luftanalysen. Senatskommission zur PruÈ fung gesundheitsschadlicher Arbeitsstoffe, 9. Lieferung. VCH, Weinheim.

Filser JG, Klein D. 2018. A physiologically based toxicokinetic model for inhaled ethylene and ethylene oxide in mouse, rat, and human. Toxicol Lett 286: 54–79.

Filser JG, Denk B, Törnqvist M, Kessler W, Ehrenberg L. 1992. Pharmacokinetics of ethylene in man; body burden with ethylene oxide and hydroxyethylation of hemoglobin due to endogenous and environmental ethylene. Archives of toxicology, 66(3), 157–163. https://doi.org/10.1007/BF01974008

International Agency for Research on Cancer (IARC). 1999. Cancer Epidemiology: Principles and Methods. World Health Organization. International Agency for Research on Cancer. Lyon, France.

Jain RB. 2020. Associations between observed concentrations of ethylene oxide in whole blood and smoking, exposure to environmental tobacco smoke, and cancers including breast cancer: Data for US children, adolescents, and adults. Environ Sci Pollut Res Int 27:20912–19.

Kirman CR, Li AA, Sheehan PJ, Bus JS, Lewis RC, Hays SM. 2021. Ethylene oxide review: Characterization of total exposure via endogenous and exogenous pathways and their implications to risk assessment and risk management. J Toxicol Environ Health Part B Crit Rev 24: 1–29.

Manjanatha MG, Shelton SD, Chen Y, Parsons BL, Myers MB, McKim KL, Gollapudi BB, Moore NP, Haber LT, Allen B, Moore MM. 2017. Dose and Temporal Evaluation of Ethylene Oxide-Induced Mutagenicity in the Lungs of Male Big Blue Mice Following Inhalation Exposure to Carcinogenic Concentrations. Environ Mol Mutagen 58:122-134.

Marsh GM, Keeton KA, Riordan AS, Best EA and Benson SM. 2019. Ethylene oxide and risk of lympho-hematopoietic cancer and breast cancer: A systematic literature review and meta-analysis. Int Arch Occup Environ Health 92(7): 919–939.

Mikoczy Z, Tinnerberg H, 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(6): 2009–2019.

Paardekooper, LM, van den Bogaart G, Kox M, Dingjan I, Neerincx AH, Bendix MB, Beest MT, Harren FJM, Risby T, Pickkers P, et al. 2017. Ethylene, an early marker of systemic inflammation in humans. Sci Rep 7: 6889. PMID: 28761087; PMCID: PMC5537290. doi:10.1038/ s41598-017-05930-9.

Recio L, Donner M, Abernethy D, Pluta L, Steen AM, Wong BA, James A, Preston RJ. 2004. In vivo mutagenicity and mutation spectrum in the bone marrow and testes of B6C3F1 lacI transgenic mice following inhalation exposure to ethylene oxide. Mutagenesis 19:215–222.

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(6): 531–539. 10.1023/a:1024891529592.

Steenland K, Stayner L, Deddens J. 2004. Mortality analyses in a cohort of 18235 ethylene oxide exposed workers: Follow up extended from 1987 to 1998. Occup Environ Med 61(1): 2–7.

Texas Commission on Environmental Quality (TCEQ). 2020. Ethylene Oxide Carcinogenic Dose-Response Assessment, CAS Registry Number: 75-21-8. Development Support Document. Final, May 15, 2020. Texas Commission on Environmental Quality (TCEQ). https://www.tceq.texas.gov/toxicology/ethylene-oxide

Törnqvist M, Osterman-Golkar S, Kautiainen A, Jensen S, Farmer PB, Ehrenberg L. 1986. Tissue doses of ethylene oxide in cigarette smokers determined from adduct levels in hemoglobin. Carcinogenesis 7 (9), 1519–152.

U.S. Environmental Protection Agency (EPA). 2005. Guidelines for Carcinogen Risk Assessment. United States Environmental Protection Agency, Risk Assessment Forum, Washington, DC.

U.S. Environmental Protection Agency (EPA). 2012. Benchmark Dose Technical Guidance EPA/100/R-12/001. U.S. Environmental Protection Agency. Risk Assessment Forum. Washington, DC 20460. June 2012.

U.S. Environmental Protection Agency (EPA). 2016a. Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide (CASRN 75-21-8): In Support of Summary Information on the Integrated Risk Information System (IRIS). EPA/635/R-16/350Fa. United States Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment. Washington, DC. https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=329730

U.S. Environmental Protection Agency (EPA). 2016b. Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide: Appendices (CASRN 75-21-8). In Support of Summary Information on the Integrated Risk Information System (IRIS). United States Environmental Protection Agency (US EPA). https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=529971&Lab=NCEA.

U.S. Environmental Protection Agency (EPA). 2022. Summary of Public Comments and Responses for the Reconsideration of the 2020 National Emission Standards for Hazardous Air Pollutants: Miscellaneous Organic Chemical Manufacturing Residual Risk and Technology Review. U. S. Environmental Protection Agency. Office of Air Quality Planning and Standards. Sector Policies and Programs Division (E-143-01). Research Triangle Park, North Carolina 27711. December 2022.

U.S. Environmental Protection Agency (EPA). 2024a. IRIS Ethylene Oxide (EtO) Assessment. Presentation to the National Academies of Sciences, Engineering, and Medicine (NASEM). Presented by Kristina Thayer and Paul White. US Environmental Protection Agency (EPA). Office

of Research and Development (ORD). Center for Public Health and Environmental Assessment (CPHEA). US EPA Chemical Pollutant Assessment Division (CPAD). June 27, 2024. https://www.nationalacademies.org/event/43049_06-2024_review-of-tceqs-ethylene-oxide-development-support-document-meeting-4

U.S. Environmental Protection Agency (EPA). 2024b. Summary of Public Comments and Responses for New Source Performance Standards for the Synthetic Organic Chemical Manufacturing Industry and National Emission Standards for Hazardous Air Pollutants for the Synthetic Organic Chemical Manufacturing Industry and Group I & II Polymers and Resins Industry. U. S. Environmental Protection Agency. Office of Air Quality Planning and Standards. Sector Policies and Programs Division (E-143-01). Research Triangle Park, North Carolina 27711. March 2024.

U.S. Environmental Protection Agency (EPA). 2024c. Summary of Public Comments and Responses for Risk and Technology Review for Ethylene Oxide Commercial Sterilization Facilities. U.S. Environmental Protection Agency. Office of Air Quality Planning and Standards. Sector Policies and Programs Division (D243-02). Research Triangle Park, North Carolina 27711. February 2024.

U.S. Environmental Protection Agency Science Advisory Board (SAB). 2015. Science Advisory Board Review of the EPAs evaluation of the inhalation carcinogenicity of ethylene oxide: Revised external review draft - August 2014 [EPA Report]. (EPA-SAB-15-012). Washington, DC: U.S. Environmental Protection Agency, Science Advisory Board.

Wilson DM, Thompson LH. 2007. Molecular mechanisms of sister chromatid exchange. Mutat Res 616:11–23.

Wu, KY, Chiang SY, Huang TH, Tseng YS, Chen YL, Kuo HW, Hsieh CL. 2004. Formation of N-(2-hydroxyethyl)valine in human hemoglobin-effect of lifestyle factors. Mutat Res 559: 73–82. PMID: 15066576. doi:10.1016/j.mrgentox.2003.12.011.