Response to Public Comments Received on the 2019 Ethylene Oxide Draft Development Support Document

CAS Registry Number: 75-21-8

Response to Comments

January 31, 2020

TEXAS COMMISSION ON ENVIRONMENTAL QUALITY

Table of Contents

List of Tables

List of Figures

Response to Public Comments Received on the June 2019 Proposed Ethylene Oxide Development Support Document

The public comment period on the draft Development Support Document (DSD) for ethylene oxide ended September 26, 2019. The agency received numerous comments on the draft DSD from diverse groups (e.g., industry, academia, nongovernmental organizations, private citizens). The TCEQ appreciates the effort put forth to provide comments on the draft DSD for ethylene oxide. The goal of the TCEQ is to protect human health and welfare based on the most scientifically-defensible approaches possible (as documented in the DSD), and evaluation of these comments furthered that goal. Comments were divided into sections and are provided below, followed by TCEQ responses.

Comments from the American Chemistry Council (ACC)

Comment 1:

The Panel supports the inhalation-based unit risk factor (URF) derived by TCEQ for ethylene oxide (EtO). TCEQ's approach to ground-truth the selection of the extrapolation model based on biological and epidemiological evidence is a critical missing step in EPA's IRIS EtO assessment (IRIS, 2016). An overly conservative assessment can result in misplaced public concern, supply chain disruption of critical products, and the unnecessary use of resources.

Response:

The TCEQ appreciates the comment. Ground-truthing and reality checks based on both biological and epidemiological evidence are rarely possible, and in this case show USEPA's URF to be biologically implausible and over-predictive while TCEQ's is both plausible and relatively accurate. Scientists must objectively review the scientific weight of evidence and the consistency or inconsistency of scientific (e.g., dose-response) assessments with it, judging each based on a detailed and unbiasedly objective review of its own merits. This duty is critically important because this and similar assessments have important regulatory, public health, and risk assessment/communication implications (e.g., whether typical environmental exposures and those near sterilization facilities represent realistic health concerns/hazards or not, and the availability of ample sterile medical supplies).

Comment 2:

The TCEQ proposed EtO DSD calculated a URF of 2.5E-6 per ppb (1.4E-6 per $\mu g/m^3$) and a 1/100,000 extra risk chronic health-based effects screening level for non-threshold dose response cancer effect of 4 ppb (7 $\mu g/m^3$) based on the NIOSH epidemiology study and an assumption of a 15-year exposure lag period. Although ACC has previously recommended a different approach based on the two strongest epidemiology studies and zero lag period, ACC finds the TCEQ proposal acceptable because it is much more scientifically sound, biologically plausible, and statistically correct compared to the IRIS (2016) EtO Assessment. The IRIS' URF of 9.1E-3 per ppb (5.0 E-3 per $\mu g/m^3$) results in a 1/100,000 excess risk concentration of 1 ppt (0.0018 $\mu g/m^3$), which is inconsistent with the epidemiological and biological evidence and

unreasonably conservative. The major reason for the 4000-fold difference in the URFs derived by TCEQ and IRIS is the selection of different statistical models used for low dose extrapolation.

Response:

Based on biological considerations discussed in the DSD, the TCEQ used a 15-year exposure lag and disagrees with the 0-year exposure lag previously proposed by ACC. Additionally, based on weighting considerations, the TCEQ ultimately based the URF on the same NIOSH key study as USEPA (2016), so ultimately disagrees with ACC on inclusion of the UCC study in the final URF. The agency agrees that the major reason for the large difference in the URFs derived by TCEQ and IRIS is the selection of different statistical models used for low-dose extrapolation. USEPA (2016) did not calculate various p-values and AIC values correctly as demonstrated in the DSD, and cannot cite MOA data that support its overall supra-linear model or the application of its very steep low-dose slope. These and other considerations such as reality checks on the underlying key cancer data and normal endogenous levels do not support adoption of USEPA's unconventional, overall supra-linear model over a standard model such as the Cox proportional hazards model, which the TCEQ shows to be more predictive and biologically plausible.

Comment 3:

TCEQ used mode of action (MoA) information as the primary basis for informing the low dose extrapolation, and systematically considered endogenous levels, key epidemiological data and model prediction to check and ground-truth the selection of the final model. Although IRIS (2016) also considered the MoA, toxicology and epidemiology studies for cancer classification, IRIS (2016) did not fully utilize these studies in the final selection of the extrapolation model. Instead, IRIS relied primarily on incorrect statistical analysis and flawed visual representation of the exposure-response data. TCEQ's approach to ground-truth the selection of the extrapolation model based on biological, epidemiological and statistical model prediction evidence is the critical missing step in the IRIS assessment that TCEQ completes in the proposed DSD.

Response:

The TCEQ agrees that the agency used multiple lines of evidence that support its approach but not USEPA's approach. The agency further agrees and has demonstrated in the DSD that USEPA relied on incorrect statistical analysis (i.e., incorrectly calculated p-values and AIC values) and an unintentional visual misrepresentation of model fit for its flawed model selection (see the DSD for a more detailed discussion).

Comment 4:

ACC has five key recommendations for strengthening TCEQ's use of mode of action and epidemiological weight of evidence to ground-truth the final selection of the URFs. While TCEQ's reality check of the EPA-estimated 1 in a million to 1 in 10,000 extra risk levels is appropriate based on endogenously generated EtO relative to those contributed by exogenous EtO exposures, it can be strengthened by brief discussion of endogenously produced EtO DNA adducts (e.g., Marsden et al. 2009, Swenberg et al. 2011, Tomkins et al. 2009).

Response:

TCEQ has chosen to now briefly highlight Recio et al. (2004) as part of the DSD MOA section as opposed to the studies referenced by the commenter. Since lymphoid cancer drove the USEPA carcinogenic assessment, perhaps the most relevant mutagenicity data discussed by USEPA (2016) was that in the bone marrow of mice exposed to EtO by inhalation *in vivo* (Recio et al. 2004), which USEPA indicates is *consistent with a linear dose-response* (see C-17 of USEPA 2016), at least at doses well above endogenous.

Comment 5:

TCEQ's arguments to support the selection of lymphoid cancer as the "critical cancer endpoint", while valid, would be enhanced by including a weight of evidence evaluation of the breast cancer findings from the six relevant epidemiology studies.

Response:

This type of information has now been included in Appendix 6 of the revised DSD.

Comment 6:

TCEQ should consider simplifying and clarifying a few sections and tables to better support TCEQ's principled approach of using MoA, biological plausibility and epidemiological weight of evidence to inform selection of the final model and the point-of-departure (PoD). The following are a couple of examples: ACC previously recommended use of zero-lag, but supports TCEQ's rationale for selecting the 15-year lag based on biological considerations and for consistency with the IRIS (2016) approach. Several tables can be simplified to only show the zero and 15-year lag data. TCEQ should clarify that the 1/100,000 extra risk level was estimated directly from the Cox proportional hazard model. This excess risk level is at the low end of the observable range of responses consistent with EPA (2005) guidance for selecting a PoD for cancer risk assessment.

Response:

Showing the results for all the various lags better documents TCEQ's extensive evaluation. Thus, the tables will not be simplified. The TCEQ has clarified in the DSD that the 1/100,000 extra risk level (the low end of the observable range) was estimated directly from the Cox proportional hazard model.

Comment 7:

ACC agrees with TCEQ's emphasis on the biological mode of action and the epidemiology weight of evidence as the primary basis for selecting the type of model for low-dose extrapolation. TCEQ also provides additional statistical evidence that the final adopted TCEQ model accurately predicts the observed number of lymphoid cancer deaths in the NIOSH cohort compared to USEPA's supra-linear spline model. Further clarifications and comparisons could be added to help the reader more fully appreciate these model-prediction results: TCEQ should clarify in Section 3.4.1.2.2.3 that regardless of whether the maximum likelihood estimate (MLE)

or the 95% upper confidence limit (UCL) model is used, the IRIS two-piece spline model over predicts the number of mortalities 95% of the time (Table 31, 95% CI). In contrast, both the MLE and the UCL for TCEQ's Cox proportional hazard log-linear model accurately predict the observed mortalities. Comparison of the prediction of the IRIS Cox proportional log-linear hazard model with the IRIS supra-linear two-piece spline model provides an additional "apples-to-apples" comparison based on similar IRIS assumptions for both model estimates.

Response:

The TCEQ appreciates the commenter's acknowledgment of the scientific weight of multiple lines of evidence approach utilized by TCEQ. To be abundantly clear, the lack of EtO-specific MOA data to justify an unconventional overall supra-linear model is the primary driver for rejection of USEPA's two-piece spline model in favor of a standard dose-response model (TCEQ 2015). However, other lines of evidence strongly support the biological plausibility and predictiveness of TCEQ's model over USEPA's model. For some additional information from Appendix 2 of the revised DSD, some sentences regarding USEPA versus TCEQ model predictions using the UCL and MLE were added to Section 3.4.1.2.2.3.

Comment 8:

TCEQ should clarify that contrary to USEPA SAB's recommendation, IRIS used only a subset of 100 randomly chosen controls from the NIOSH data (IRIS Appendix D-4, D-29), whereas, TCEQ's model estimates are based on the full NIOSH data set.

Response:

Text was added to the DSD to the effect that TCEQ lymphoid cancer risk estimates are based on the full NIOSH and UCC datasets (i.e., the individual data and not categorical results). Although USEPA IRIS used a random sample of 100 individuals in defining risks sets in the process of fitting the models, results are close to using all individuals. TCEQ used all individuals to define the risk sets, which makes a more robust set for estimating model parameters. This response to comments document serves as documentation of the comment that contrary to USEPA SAB's recommendation, IRIS used only a subset of 100 randomly chosen controls from the NIOSH data.

Comment 9:

TCEQ appropriately relies on the biological MoA as the primary basis for selecting the model for low-dose extrapolation to build a strong case for why TCEQ should not adopt the EtO IRIS assessment's inhalation of 1 in 100,000 excess risk-based air concentration of 1 ppt. TCEQ's conservative and scientifically supportable approach to an exposure-response analysis should be used. This alternative approach makes use of the full data set and yields a more realistic risk-based air concentration of 4 ppb at the no significant excess risk level of 1 in 100,000.

Response:

The TCEQ agrees that biological MOA should be the primary basis for selecting the dose-response model; the scientific weight of evidence strongly supports TCEQ's approach over

USEPA's modeling approach; and that TCEQ's approach incorporates appropriate conservatism yet remains realistic in terms of biological plausibility and risk predictiveness.

Comment 10:

p. 60- last paragraph regarding Table 10. TCEQ states that "NIOSH breast cancer incidence data were not publicly available for independent analysis. Therefore, Table 10 results will not be utilized." Perhaps these two sentences can be switched.

Response:

The TCEQ has clarified the referenced text in the DSD.

Comment 11:

p. 64- first sentence in italics explains the rationale for excluding breast cancer as a final endpoint. This section should incorporate consideration of the weight of evidence for breast cancer incidence. The epidemiology data does not support a potency for breast cancer that is stronger than for lymphoid cancer.

Response:

Relevant information has been included in Appendix 6 of the revised DSD.

Comment 12:

p. 84 and 90- the statement is made in reference to Swaen et al. (2009) and Mikoczy et al. (2011) that "Healthy Worker Effect (HWE)" likely influenced results". HWE is a well-known form of bias in occupational cohort studies in which increased risks may be missed when comparisons are made to an external, general population, considered to be less healthy than the worker population. However, the epidemiologic literature has shown that HWE is predominately related to shorter follow up and non-cancer causes (Monson 1986; Fox and Collier 1976). Swaen (2009) had a very long follow up (36.5 yr. average) and deficits in major non-cancer causes only among those hired after 1956. There is no indication that cancer increases have been missed due to HWE. Similarly, for Mikoczy et al. (2011), mortality was no longer decreased with a 15-yr "induction latency" period. A study to test HWE in Sweden as it relates to breast cancer has been published showing no HWE (Gridley et al. 1999). To avoid misleading the reader, we recommend deleting these statements in the report or specifying that they relate to non-cancer causes.

Response:

The TCEQ agrees that the healthy worker effect (as evidenced by decreased overall mortality, etc.) does not necessarily extend to specific cancers, particularly where the carcinogen operates via a mutagenic MOA. For example, the suggestion of the authors of Mikoczy et al. (2011) that a finding of significantly decreased overall mortality and cardiovascular disease mortality is indicative of a healthy worker effect for breast cancer incidence is inconsistent with the results of a relatively recent and large study (366,114 workers) conducted specifically to examine the potential for the healthy worker effect in cancer incidence studies (Kirkeleit et al. 2013).

Similarly, the authors of Swaen et al. (2009) indicated that results for all-cause mortality, heart disease, non-malignant respiratory diseases, etc. indicated a healthy worker effect necessitating internal analyses for cancers of interest. In Kirkeleit et al. (2013), all-cause mortality and both ischemic heart disease and circulatory system disease mortality were statistically significantly decreased in male workers (n=283,002) and female workers (n=83,112) compared to the general population (Table 3 of the study), consistent with similar findings in Mikoczy et al. (2011) and Swaen et al. (2009). The SIRs for lymphoid and hematopoietic cancers in male workers and female workers were 0.97 (0.90, 1.03) and 1.09 (0.92, 1.27), respectively, consistent with the lack of a statistical difference; that is, the lack of a healthy worker effect for these cancers. Additionally, the Kirkeleit et al. (2013) study found that breast cancer incidence in over 83,000 female workers was as expected based on the general population (i.e., SIR of 1.02 (0.95, 1.09)). Thus, Kirkeleit et al. (2013) indicates the lack of a healthy worker effect for both lymphoid/hematopoietic and breast cancer incidence. In any event, both Mikoczy et al. (2011) and Swaen et al. (2009) conducted internal analyses to account for the presumed healthy worker effect, so these study results need not be caveated. Therefore, the referenced text was removed as the study authors accounted for the presumed healthy worker effect for cancers of interest.

Comment 13:

TCEQ should consider simplifying and clarifying a few sections and tables to better support TCEQ's principled approach of using MoA, biological plausibility and epidemiological weight of evidence to inform selection of the final model and the point-of departure (PoD).

Table 6 (p. 56) includes some cancer endpoints that are not relevant based on the epidemiological weight of evidence. This table should only include lymphohematopoietic and breast cancers, which are the only cancers that IRIS (2016, p. 3-13) associated with EtO exposures.

Response:

The TCEQ understands this perspective but does not perceive a need to change the table as it helps demonstrate the broader nature of analyses conducted for EtO.

Comment 14:

Table 7-10, 12-14 (pp. 57-62) can be simplified to just show the zero and 15-year lag. TCEQ should indicate in the text and footnote of these tables that a large number of lag periods were tested and none were statistically different from zero lag. ACC previously recommended use of zero-lag, but supports TCEQ's rationale for selecting the 15-year lag based on biological considerations and for consistency with IRIS (2016) approach. However, it should be noted that in some cases the 95% UCL URFs for zero lag were slightly higher (more conservative) than for the 15-year lag.

Response:

The various exposure lags in Table 7-10 help demonstrate the broader nature of analyses conducted for EtO. The TCEQ did add language to the text and footnotes indicating that no lag duration fit statistically better than the model with zero lag. The other points made by the commenter are duly noted, including support for TCEQ's rationale for selecting the 15-year lag based on biological considerations and for consistency with IRIS (2016) approach.

Comment 15:

Tables 12-14 should add explanations that the 1/100,000 extra risk level was estimated directly from the Cox proportional hazard model, and that this is consistent with EPA (2005) cancer guidelines on selection of the PoD at the low end of the observable range of responses. For example, with rodent models, a 10% (1 in 10) PoD is typically used as a 10% extra risk and is near the limit of detection for a typical assay. For epidemiologic data, a lower PoD can be used. When the standard Cox proportional hazard (log-linear) model is used for the NIOSH males-only 15-year lag data, all of the lymphoid mortalities with non-zero exposure occurred below the 1 in 100 PoD (Table 2). Therefore, 1 in 100 is not an appropriate PoD for "extrapolation" in the conventional sense. A typical POD extrapolates from the edge of the observed range through the unobserved range of the data. Thus, for the NIOSH male only data, it is appropriate to use the model to extrapolate to 1 in 100,000, which is below the 50th percentile of exposure where there is only one lymphoid mortality for subjects with non-zero exposure. IRIS (2016) used a 1% (1 in 100) extra risk for the PoD but did not provide evidence that this level would establish a PoD near the edge of the observed data range. ACC does not have the NIOSH data to determine the validity of the 1% for the supra-linear spline model.

Table 2. Number of male lymphoid cases out of approximately 18,000 workers with concentrations below the EC (1/100) and EC (1/100,000)

	Male Lymphoid EC 1/100		Male Lymphoid EC 1/100,000[2]	
	0-Lag	15-Lag	0-Lag	15-Lag
EC (1/100,000) Env. Conc (ppm)	3.52	5.80	5.83E-03	9.67E-03
Equivalent[1] Occupational Exposure 70 years (ppm-days)	326,105.9[2]	354,399.0[2]	453.4[2]	590.87[2]
Total Number of Deaths	27	27	27	27
Number with zero exposure	0	6	0	6
Number With Non-Zero Exposure below EC	27	21	1	1
Percentage of Deaths below EC	100%	100%	3.70%	25.93%

[1]Equivalent Occupational Exposure 70 years (ppm-days) = EC×(365/240)×(20/10)×365.25×(70-lag)

[2]The maximum occupational exposure concentration for lymphoid deaths was less than 326,106 ppm-days for the unlagged and 137,243 ppm-days for the 15-year lag exposure.

Response:

The relevant tables have been footnoted accordingly. TCEQ evaluated the LEC at an extra risk of 1 in a 100,000 consistent with USEPA cancer guidelines (2005) on the selection of a POD at the low-end of the observable range of exposures. Although for animal studies, a typical POD is an extra risk of 0.10 because it corresponds to doses near the low-end of the doses, in epidemiological studies a lower level of risk needs to be used.

TCEQ used the standard Cox proportional hazards model to calculate LEC for an extra risk of 1 in a 100,000 because the EC corresponding to the same risk level are in the range of the observed data in the NIOSH study. That is, the EC for an extra risk of 1 in 100,000 of lymphoid cancer mortality in males is 9.67E-03 ppm for 70 years with an exposure lag of 15 years, which correspond to a cumulative occupational exposure of

591 ppm-days. There are 7 male workers in the NIOSH cohort with cumulative exposures less than 591 ppm-days. That is, 25.9% of the male workers in the NIOSH cohort that died with lymphoid cancer were exposed to cumulative exposures of less than the EC for 1 in a 100,000. In contrast, the EC for 1 in 100 results in environmental concentrations corresponding to cumulative occupational exposures of 354,400 ppm-days, which exceeds the largest cumulative exposure of lymphoid male decedents in the NIOSH study.

The following table shows the EC corresponding to different risk levels and the corresponding cumulative exposures with the number of lymphoid mortality cases of the male workers in the NIOSH study.

Table 1. Environmental and equivalent occupational cumulative exposures for different potential points of departure using TCEQ's selected model for lymphoid mortality in the NIOSH study

Statistic	Extra Risk			
	1/100	**1/1,000**	**1/10,000**	**1/100,000**
Environmental EC (ppm) [1]	5.80×10^{-0}	8.99×10^{-1}	9.61×10^{-2}	9.67×10^{-3}
Equivalent Occupational EC (ppm-days) [2]	354,399	54,932	5,872	591
Lymphoid Deaths [3]	27	21	13	7
% Lymphoid Deaths [4]	100%	77.78%	48.15%	25.93%
% Male Workers [5]	99.84%	94.48%	66.45%	30.17%
LEC (ppm) [6]	2.44×10^{-0}	3.78×10^{-1}	4.04×10^{-2}	4.07×10^{-3}
URF (ppb^{-1}) [7]	4.09×10^{-6}	2.64×10^{-6}	2.47×10^{-6}	2.46×10^{-6}

[1] Environmental concentration in ppm for 70-year lifetime with lag of 15 years corresponding to a specified extra risk
[2] Equivalent Occupational Exposure 70 years (ppm-days) = EC (ppm) × (365/240 days) × (20/10 m^3) × (365.25 days/year) × (70 years – lag in years)
[3] Number of male workers in the NIOSH cohort that died of lymphoid cancer with cumulative exposure less than the EC (i.e., EC in ppm-days at 1/100, 1/1,000, 1,10,000, or 1/100,000)
[4] Percentage of lymphoid cancer decedent male workers in the NIOSH cohort with cumulative exposures less than the EC (ppm-days)
[5] Percentage of male workers in the NIOSH cohort with cumulative exposures less than the EC (ppm-days)
[6] 95% lower bound on the EC (ppm)
[7] Unit risk estimate based on the LEC (ppm)

The results in Table 1 show that the EC for an extra risk of 1 in a 100 is outside the range of cumulative exposures for the male lymphoid mortalities observed in the NIOSH study and in the upper 1% of cumulative exposures for all male workers. That is, all males that died with lymphoid cancers and more than 99% of all male workers had cumulative exposures less than EC(1/100). Thus, the NIOSH study does not support an extra risk of 1 in a 100 as a point of departure.

The EC for an extra risk of 1 in a 1,000 is a concentration that is in the high-end of cumulative exposures of male lymphoid mortalities observed in the NIOSH study. That

is, 77.78% of all males that died with lymphoid cancers and 94.48% of all male workers had cumulative exposures less than the EC(1/1,000). Thus, a point of departure of 1 in 1,000 is at the higher-end of the cumulative exposures of male workers of the NIOSH study.

The EC for an extra risk of 1 in 10,000 is a concentration that includes 48.15% of the decedent men with lymphoid cancer and 66.45% of all men in the NIOSH cohort with smaller cumulative exposures. *The EC for an extra risk of 1 in 100,000 includes 25.93% of male lymphoid decedents and 30.17% of all males in the NIOSH study with smaller cumulative exposures.* Thus, use of an extra risk of 1 in 100,000 is supported by the NIOSH observed data, being near the lower end of the observed range of cumulative exposures to EtO, and is consistent with USEPA and TCEQ guidelines (USEPA 2005a, TCEQ 2015) on the selection of a POD at the low-end of the observable range of exposures.

Based on Table 1 results, using either 1 in 10,000 or 1 in 100,000 extra risk PODs (as PODs in the range of the observed data and close to the low-end of the observable range) round to the same ADAF-unadjusted URF selected by the TCEQ (2.5E-06 per ppb). Looking at it from a different perspective, using the 1 in 10,000 excess risk LEC of 4.04E-02 ppm as the POD and linear extrapolation, the 1 in 100,000 air concentration (ADAF unadjusted) is still 4 ppb (i.e., 1E-05/2.47E-06 per ppb = 4.05 ppb). This information has been added to Appendix 7 of the revised DSD.

Comment 16:

ACC agrees with TCEQ's emphasis on the biological mode of action and the epidemiology weight of evidence as the primary basis for selecting the type of model for low-dose extrapolation. TCEQ also provides additional statistical evidence that the final adopted TCEQ model accurately predicts the observed number of lymphoid cancer deaths in the NIOSH cohort compared to EPA's supra-linear spline model.

Response:

The agency acknowledges the commenters concurrence with TCEQ's approach. As indicated, the model adopted by the TCEQ is demonstrated to relatively accurately predict the observed number of lymphoid cancer deaths in the NIOSH cohort compared to the over-prediction by USEPA's supra-linear spline model.

Comment 17:

Further clarifications and comparisons could be added to help the reader more fully appreciate model-prediction results.

P. 41-46, Section 3.4.1.2.2.3: TCEQ used the final selected 95% upper confidence limit (UCL) model to predict lymphoid mortalities. TCEQ may want to further clarify that regardless of whether the maximum likelihood estimate (MLE) or the 95% upper confidence limit (UCL)

model is used, the IRIS two-piece spline model over predicts the number of mortalities 95% of the time (Table 31, 95% CI).

Response:

This is now mentioned at the end of Section 3.4.1.2.2.3.

Comment 18:

In contrast, the MLE and UCL models for TCEQ's Cox proportional log-linear model accurately predicts the number of mortalities. The section on model prediction analysis could also clarify that this comparison is based on the model fit prior to any additional adjustments based on age or other factors.

Response:

The agency believes no such clarification is needed as it is well before the ADAF section of the DSD.

Comment 19:

Figures 8 to 12: TCEQ might consider including IRIS's Cox proportional log-linear model in Figures 8 to 12 for comparison with IRIS's supra-linear two-piece spline slope. Comparison of the prediction of the IRIS Cox proportional log-linear hazard model with the IRIS supra-linear two-piece spline model provides an additional comparison based on similar IRIS approach (i.e. using a random subset of the data).

Response:

While considered, the agency believes no such revision is needed since the TCEQ is using its own Cox proportional hazarding modeling results for comparison to USEPA's selected model.

Comment 20:

TCEQ should clarify that contrary to USEPA SAB's recommendation, IRIS used only a subset of 100 randomly chosen controls from the NIOSH data (IRIS Appendix D-4, D-29), whereas, TCEQ's model estimates are based on the full NIOSH data set.

USEPA SAB recommended that IRIS utilize the full NIOSH data set to estimate the cancer slope coefficients that would in turn be used to extrapolate risk instead of a small subset used by IRIS (IRIS Appendix H-10).

TCEQ's model estimates are based on the full NIOSH data set. However, the IRIS (2016) model use the subset of 100 controls. There is no strong biologic or statistical justification for selecting a subset of the data to estimate dose response curves. Thus, TCEQ's analysis is a more robust and complete analysis based on all the available data.

Response:

See TCEQ response to Comment #8 (above). As noted in the next comment, TCEQ's model estimates are based on the full NIOSH dataset, consistent with the SAB recommendation. The agency acknowledges the commenter's concurrence with our approach.

Comment 21:

p. 14 and p.27 authorship should be corrected in the section in italics regarding update of the UCC cohort. Dr. Valdez-Flores is not a co-author of the Bender et al. 2019 paper (submitted), but is an author of a risk assessment paper based, in part, on the Bender et al. 2019 paper.

Response:

This clarification has been made.

Comment 22:

p.25, para.2: This text effectively describes how the implausibly high cancer risk associated with low dose EtO exposures as estimated by USEPA also infers an implausibly high cancer risk associated with exogenous long-term exposure to ambient levels of ethylene (due to its metabolism to EtO). However, the analysis should be expanded to clarify that, unlike EtO, the current risk assessments for ethylene are based on robust negative chronic rodent inhalation bioassays and genotoxicity assessments, and thus should not be targeted for cancer risk reevaluation based on extrapolation from the USEPA EtO cancer risk assessment.

Response:

Some clarifying language has been added to Section 3.4.1.2.1 of the DSD.

Comment 23:

Table 4 A footnote should be added next to Valdez-Flores et al. 2010 that only the first and fourth column are based on data from Valdez Flores et al. 2010.

Table 4 The breast cancer row incorrectly indicates the highest 5th quantile is elevated risk, but we believe this is incorrect because there was no statistical increase. Instead it should indicate Not Applicable.

Table 5 Similar to Table 4, a footnote should be added to clarify that only columns 1 and 4 are from Steenland et al. (2004, 2003).

Response:

The recommended footnote has been added to the referenced tables. Breast cancer was included in Table 4 of the proposed DSD by TCEQ for context, and column 1 denotes that it was not statistically increased. This caveat was added to the fourth column of this table.

Comment 24:

p. 57-60 This series of tables was difficult to follow. We recommend separating the p-value vs. null and p-value vs. zero lag into separate columns by themselves.

Response:

No changes were deemed necessary as the tables fulfill the purpose of adequately documenting the information contained therein.

Comments from Sierra Club, Texas Environmental Justice Advocacy Services, Air Alliance Houston, Coastal Alliance to Protect our Environment, Environment Texas, Public Citizen's Texas Office, Texas Campaign for the Environment, Earthjustice, and Environmental Integrity Project

In their letter to the agency, these commenters provided many general comments on ethylene oxide (EtO) as well as more specific comments relevant to the TCEQ carcinogenic dose-response assessment under consideration, which were assembled by the TCEQ and are considered below.

Comment 1:

EtO is a well-known human carcinogen. Children are particularly vulnerable to mutagenic carcinogens and exposure during early life further increases the likelihood of developing cancer.

Response:

Indeed, the TCEQ's assessment treats EtO as a mutagen that is *carcinogenic to humans* (i.e., a known human carcinogen). Consistent with USEPA guidance, the TCEQ uses age-dependent adjustment factors (ADAFs) to account for the potentially increased susceptibility of children. Additionally, the TCEQ uses the 95% UCL to err on the side of safety and help ensure that vulnerable people and communities are protected against the potential carcinogenic effects of EtO in ambient air (although normal endogenous doses produced within the human body are significantly higher than those resulting from typical ambient air concentrations).

Comment 2:

Due to the serious cancer risk from exposure to EtO, in 2016, USEPA completed a robust, scientific, and peer-reviewed process to protect public health and finalize a toxicity factor for EtO of 0.005 per $\mu g/m^3$, or 0.0091 per ppb. USEPA demonstrated that breathing just 0.0002 of a microgram of ethylene oxide per cubic meter of air, or 0.0001 parts ethylene oxide per billion parts air over a lifetime increases cancer risk by 1-in-1 million. EPA's cancer risk factor is "based on strong epidemiological evidence supplemented by other lines of evidence" on lymphoid and breast cancers, and accounts for the increased risk to children through applying age-adjustment factors. USEPA has "relatively high" confidence in its factor as an estimate of the upper bound on risk from lifetime exposure, with "particularly high" confidence for its breast cancer component.

Response:

The TCEQ has scientifically demonstrated in this DSD that USEPA's selected model assessment statistically significantly overestimates risk for workers in the underlying key study, whereas the TCEQ's model accurately predicts risk. Furthermore, USEPA has not demonstrated that breathing just 0.0002 $\mu g/m^3$ (or 0.0001 ppb) over a lifetime increases cancer risk by 1-in-1 million. This is a theoretical risk estimate based on an atypical dose-response model that was

selected using incorrect p-values and AIC values (see Appendices 3 and 5 of the DSD). As discussed in other comment responses, the TCEQ also uses ADAFs to account for the potentially increased susceptibility of children. While USEPA had high confidence in its URF, the TCEQ has demonstrated it to overestimate risk for both workers in the underlying key study and the US general population. These demonstrations in conjunction with other information, particularly the USEPA-acknowledged lack of MOA information to justify their unconventional overall supra-linear dose response model, significantly diminished the TCEQ's confidence in USEPA's dose-response assessment choices and URF and prompted the TCEQ to conduct its own dose-response assessment. See Appendix 6 of the revised DSD, which is new, for a detailed discussion on the evidence for breast cancer as a candidate endpoint.

Comment 3:

Since USEPA finalized the IRIS factor for EtO in 2016, TCEQ has made every effort to ignore or discredit it. TCEQ appears to favor industry in pushing weaker and weaker factors that fail to protect Texas communities - especially those already overburdened by toxic air pollution, and particularly women and children.

Response:

In the 2019 EtO DSD, the TCEQ documents the completion of an objective and thorough scientific evaluation of USEPA's assessment, which concluded that the USEPA's URF is not scientifically sound. For example, reality checks conducted by the TCEQ on risk in the key worker population and the general population could have supported USEPA's URF (the outcome was unknown), but simply did not. The USEPA's assessment must stand or fall on the objective evaluation of its merit, and the TCEQ has thoroughly documented its objective evaluation of same. The TCEQ's goal is to have the most scientifically defensible URF possible, and the TCEQ will follow wherever the scientific weight-of-evidence leads. The TCEQ's only effort and motivation is to be on the side of the best available science in this case and in all the agency does to protect public health and the environment. The TCEQ's EtO URF is protective for all people, including sensitive populations, wherever they reside in Texas.

Comment 4:

In March 2017, TCEQ adopted a risk factor of 0.000076 per $\mu g/m^3$ - 65 times weaker than the 2016 IRIS factor. In a three-page document announcing the 2017 factor (that only became available after a Public Information Request to TCEQ by Sierra Club), TCEQ considered and rejected two studies - just as EPA had done: Valdez-Flores et al. (2010) and Kirman et al. (2004). TCEQ rejected Valdez-Flores et al. (2010) because it failed to capture cancer risk for all but the highest exposure groups, and rejected Kirman et al. (2004) for various reasons, including its failure to consider breast cancer. The value selected in 2017 by TCEQ was actually part of USEPA's IRIS assessment but based just on rodent data (as opposed to the final 2016 IRIS risk factor which was based on the entire systematic review, including human data). TCEQ selected the 2017 factor due to the "high quality" of the rodent study, without providing a reasoned basis then for rejecting the remaining conclusions of USEPA's determination or the final 2016 IRIS factor.

Subsequently, in August 2017, without reference to TCEQ's March 2017 factor or conclusions, without any explanation, and without any apparent reason, TCEQ began to create a new factor that was even weaker. TCEQ publicized a request for information on its website, and the American Chemistry Council (ACC) submitted comments urging TCEQ to develop a weaker factor, like that of Valdez-Flores et al. (2010). TCEQ was close to releasing its proposed assessment in June 2018 when it met with the ACC.

Response:

The TCEQ performs an expedited review of USEPA IRIS and other toxicity factors when they are promulgated and may adopt them or an alternative value on an interim basis. More specifically, the TCEQ adopted an animal-based URF for EtO on an interim basis (recorded with a 1.5-page document) until a thorough dose-response assessment could be conducted under the extensive TCEQ toxicity factor guidelines and a detailed discussion of scientific rationale could be documented in the DSD. A much longer period of time is required for a systematic review of the peer-reviewed literature and for producing detailed DSD documentation, including a thorough evaluation of the USEPA's selected model and URF in this case. A request for EtO information was put out by TCEQ later in the year following the interim value derivation, as it was obvious to TCEQ that a more detailed review was in order. In response to the TCEQ request for EtO information, *all* interested parties (e.g., industry, academia, NGOs, private citizens) were invited to submit whatever they believed to be relevant information (e.g., studies, presentations) for the agency's evaluation, to be considered along with the information that the TCEQ identified through an extensive systematic review process for EtO. In June 2018, the TCEQ was still in the early stages of conducting our systematic review and none of the extensive analyses had been conducted.

Comment 5:

Just as instructed by the ACC, and despite TCEQ's own March 2017 conclusions, TCEQ (1) selected Valdez-Flores et al. (2010) as its key study; (2) incorporated the unpublished, not peer-reviewed update, Bender et al., with the help of Dr. Valdez-Flores, an EtO and sterilant trade group consultant; (3) ignored breast cancer, even though TCEQ admits that breast cancer incidence data supports a much stronger toxicity factor, because the "results [were] not consistent with TCEQ conclusions;" and (4) ignored what it described as endogenous exposure. Neither TCEQ's proposed DSD, Dr. Valdez-Flores's analyses, nor the underlying study Bender et al. have undergone any independent peer review. And, the study measuring endogenous exposure suggests normal, endogenous levels of EtO more than 65 times higher than the equivalent exogenous exposure of living directly next to a sterilizer facility, like Willowbrook or Burr Ridge.

Response:

The TCEQ's derivation of toxicity factors is instructed by the best available science and risk assessment methodologies, as documented in the proposed EtO DSD. The revised DSD and accompanying analyses will be undergoing peer review in the first quarter of 2020. To the individual points in this comment:

1) The TCEQ conducted a systematic review of the literature (documented in Appendix 1 of the DSD), and based on this selected the methods used by Valdez-Flores et al. (2010) as the most scientifically valid. However, the agency did not use the results proposed by Valdez-Flores et al. (2010) to derive the final URF, but rather used their method with alternative model parameters that were dictated by the TCEQ.

2) Dr. Valdez-Flores is a Professor at Texas A&M University and has conducted excellent statistical work (as directed) for the TCEQ previously. The agency has found his expertise to be outstanding and his integrity to be beyond reproach. The TCEQ did consider the updated UCC cohort data (Bender et al.); however, data from this cohort ultimately received a zero weight for the final URF because this dataset was substantially smaller than the NIOSH cohort.

3) The TCEQ evaluated breast cancer as a candidate cancer endpoint. USEPA's assessment, however, is driven by lymphoid cancer as the primary contributor to the URF. Subsequent to USEPA's 2016 assessment and TCEQ's systematic review of the peer-reviewed literature for EtO, recent meta-analyses of available studies have been published (Marsh et al. 2019, Vincent et al. 2019). These new meta-analyses included Steenland et al. (2003, 2004) and the smaller Mikoczy et al. (2011) study cited by USEPA (2016), as well as other studies, and reported breast cancer meta-RRs of 0.97 (0.80, 1.18) (Marsh et al. 2019) and 0.92 (0.84, 1.02) (Vincent et al. 2019). The Marsh et al. study concluded [*emphasis added*], "Evaluations of workers exposed during sterilization processes *do not support the conclusion that EO exposure is associated with an increased risk of breast cancer*." Similarly, the Vincent et al. (2019) study concluded, "Higher quality epidemiological studies demonstrated no increased risk of breast cancers." In addition to evaluating epidemiological evidence, Vincent et al. (2019) evaluated animal study results and concluded that they provide no strong indication that EtO causes mammary tumors. These recent meta-analyses and other information (IARC 2019) further support TCEQ's decision not to base the URF on breast cancer. Human data are by far the most relevant for derivation of human toxicity factors, and the human data themselves are inconclusive (as acknowledged by USEPA 2016). The weight of evidence for EtO-induced breast cancer is now discussed in a new appendix (Appendix 6) to the revised DSD.

4) The TCEQ extensively evaluated endogenous EtO exposure and used it for context in the DSD. For example, the endogenous data show that normal endogenous doses are orders of magnitude higher than that associated with USEPA's maximum acceptable risk-based air concentration. The TCEQ's initially-proposed long-term ESL (4 ppb) and new revised, ADAF-adjusted ESL (2.4 ppb; adjusted per equation 5-17 of TCEQ 2015) correspond to internal doses within the range of EtO doses normal produced endogenously in the human body.

Comment 6:

The results of TCEQ's assessment appear predetermined. Industry wanted a weaker factor, and TCEQ is giving it to them. Far from rational and reasoned decision-making, TCEQ's attack on IRIS was and is merely a means to that end. For example, TCEQ had apparently chosen its factor before it decided how far off to allege the IRIS factor was, at one point claiming the IRIS factor overestimated 1,179 deaths, later revised to just 141. TCEQ still has refused to release the studies and calculations it relies on in its proposed DSD, such as Bender et al. Commenters and the public still have not been provided a reasonable opportunity to evaluate the basis for TCEQ's proposed DSD.

Response:

The TCEQ has conducted extensive analyses of multiple lines of evidence, each of which could have either supported a given approach for the EtO dose-response assessment, or not. The results of each set of analyses have guided TCEQ's decisions and approach, as is documented in the DSD. The proposed DSD does not indicate that the IRIS factor overestimates by 1,179 deaths because that figure was an interim placeholder figure in early draft documents while information about the relationship between the upper and lower splines of their two-piece spline model could be located in USEPA's documentation. The proposed TCEQ URF does not and would not rely on the unpublished Bender et al. study (which will likely be published in the peer-reviewed literature in 2020). TCEQ initially provided a 45-day review period, which was extended by an additional 45 days when Sierra Club requested more time. The TCEQ thinks that the 90-day public comment period on the DSD provided interested knowledgeable parties with sufficient time to formulate scientific comments, based on the extensive scientific comments received by the TCEQ.

As part of the proposed DSD, the TCEQ commissioned outside research on the subject. This research was conducted by a professor at Texas A&M University (Dr. Ciriaco Valdez-Flores), who had previously obtained access to and extensively analyzed NIOSH data that was not readily available to TCEQ researchers, but had also been analyzed by USEPA and others (e.g., Steenland et al. 2004). Much of this information consisted of Personal Identifiable Information (PII) of a large number of research subjects. PII is information that could potentially identify a specific person, including such sensitive information as birth dates, home addresses, phone numbers, and assorted other information (e.g., death dates, gender, race, job history, cause of death). This type of information is protected by law in Texas and cannot be released without specific reasons that do not exist here. TCEQ is serious about protecting PII. Consequently, TCEQ does not have the data to release to the public in this case, although USEPA and others have analyzed the same data (see results in USEPA 2016) using the model selected by TCEQ (Cox proportional hazards model) and Sierra Club can request access to the data from NIOSH. Indeed, USEPA recently proposed a rule that would require the release of PII used in scientific research such as at issue here. Sierra Club submitted comments on that rulemaking objecting that the release of such information was inadvisable, as it could potentially cause harm to those involved in such research. Thus, even if TCEQ had the data in its possession, the agency's

refusal to release such data to Sierra Club in this case would be both required by law and the appropriate way to protect those who participate as subjects in necessary scientific research.

Comment 7:

TCEQ's proposed value is not peer reviewed, ignores breast cancer risk and ignores increased risk to children. It also treats EtO that people breathe as equivalent to endogenous exposure that TCEQ argues can be ignored.

Response:

The TCEQ assessment has undergone internal peer review and public comment but will also undergo an external peer review in the first quarter of 2020. See the response to Comment #5 and Appendix 6 of the revised DSD concerning how the agency has evaluated breast cancer as a candidate cancer endpoint. The TCEQ has used ADAFs to account for the potentially increased susceptibility of children. Lastly, as discussed in response to Comment #5, the TCEQ carefully considers and quantifies the risk posed by endogenous EtO and uses it to provide context for the EtO URF. Furthermore, the dose-response for workers exposed to EtO in the cohort studies inherently captures endogenous EtO risk (a contributor to background) as the workers were also producing EtO endogenously in addition to being exposed to exogenous EtO via the inhalation pathway.

Comment 8:

TCEQ may not finalize the proposed DSD because it has not satisfied the rulemaking requirements of Texas law. TCEQ's proposed DSD is part of an unlawful rulemaking. TCEQ's proposed DSD is a major environmental rule, requiring an in-depth regulatory analysis and impacts.

Response:

TCEQ ESL values are not developed via rulemaking. ESLs are not rules, and the process of developing an ESL is not a rulemaking action subject to the Texas Administrative Procedures Act (APA); the ESL is also not a major environmental rule. TCEQ has been developing and using ESLs for decades, and that process has never been subject to the requirements of the Texas APA. The fact that TCEQ has solicited comments on the proposed ESLs and sought input from the public does not change the essential nature of the proposed ESLs. Seeking further information and public input when developing guidance ensures that TCEQ has the broadest possible set of information to use when evaluating possible changes to the proposed guidance.

ESLs are constituent-specific guideline concentrations used in TCEQ's effects evaluation of constituent concentrations in air. These guidelines are derived by the Toxicology, Risk Assessment, and Research Division and are based on a constituent's potential to cause adverse health effects, odor nuisances, and/or effects on vegetation. Health-based screening levels are set to protect the general public, including sensitive subgroups such as children, the elderly, or people with existing respiratory conditions. They are intended to be an aid in the process of reviewing permit applications for air quality permits. The use of an ESL simply assists the

permitting authority in determining if a more extensive review of potential health effects is necessary for particular proposed changes that may result in emissions of air contaminants. ESLs, expressed in terms of microgram per cubic meter ($\mu g/m^3$) or parts per billion by volume (ppb_v) in air, are used to evaluate the potential for effects to occur as a result of exposure to concentrations of constituents in the air. Again, ESLs are based on data concerning health effects, odor/nuisance potential, and/or effects on vegetation. They are not ambient air standards.

Lastly, the proposed guidelines will only potentially affect a small subset of plants in Texas. A search of TCEQ's emissions inventory shows between 27 and 50 plants that might potentially be subject to the guidance of the Ethylene Oxide ESL, of which only 11 had reportable quantities of emissions in 2017. The ESL guidance would only affect these plants should they make permitting changes to the facilities that produce or handle ethylene oxide after the proposed guidance is finalized.

Comment 9:

TCEQ guidelines and regulations require TCEQ to adopt the IRIS cancer risk factor for EtO. TCEQ guidelines and regulations direct use of the IRIS factor and TCEQ has given no valid basis for rejecting the 2016 IRIS factor. TCEQ cannot finalize the proposed cancer risk factor because TCEQ has not provided any valid reason to develop its own cancer risk factor.

Response:

TCEQ guidelines and regulations do not require TCEQ to adopt the IRIS cancer factor for EtO. The TCEQ has evaluated the USEPA assessment extensively and has determined, as documented in the DSD, that the USEPA URF is scientifically unsupportable; a valid basis for its rejection. For example, neither agency can cite MOA information for EtO supporting USEPA's overall supra-linear dose-response model, alone a valid reason under TCEQ guidelines for rejecting the model (other considerations are also detailed in the DSD).

Comment 10:

TCEQ claims that it has performed a "reality check" of the exposure-response model selected by USEPA and found that the observed cancer deaths would be higher if USEPA's value were accurate. Its argument has no rational basis in the record and does not provide a reasoned ground to refuse to apply the IRIS value.

Response:

A demonstration that a dose-response model is not predictive of the underlying data is a rational reason for rejection of that model. As documented in the DSD, the TCEQ also found that the USEPA URF is over-predictive for lymphoid cancer in the general population based on data about background levels of EtO. The primary deterministic factor for the rejection of USEPA's model under TCEQ guidelines is the USEPA-acknowledged lack of MOA data for EtO supporting an overall supra-linear dose-response curve. The scientific burden of proof for choosing a supra-linear dose-response is robust MOA data (including data over the endogenous

range where both agencies expect sublinearity). This burden of proof was not met. Relevant considerations (including MOA) support TCEQ's approach as opposed to USEPA's, as comprehensively documented in the DSD.

Comment 11:

TCEQ must use a supra-linear modeling approach; its discussion of "endogenous exposure" does not justify using, in effect, a threshold approach. USEPA properly applied a supra-linear model to assess cancer risk from EtO. TCEQ proposes, however, that the carcinogenicity of EtO "is no more than linear, and arguably sublinear."

Response:

The scientific burden of proof for choosing a supra-linear dose-response is robust MOA data. The USEPA's 2016 EtO assessment stated, and TCEQ agrees, that "EPA 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." This information does not support a supra-linear dose-response for air EtO concentrations that are in the range of endogenous concentrations (i.e. typical ambient air EtO concentrations). The TCEQ has demonstrated USEPA's supra-linear modeling approach to be: (1) over-predictive for the key cohort lymphoid cancer data as well as the general population; (2) unjustified due to the lack of MOA data to adequately support it under TCEQ guidelines (USEPA also acknowledges the lack of MOA data to justify their model); and (3) despite the above, applied by USEPA over the very range where both agencies would expect a sublinear dose-response. These considerations are important since the occupational cohort data (used to derive the dose-response model) are at daily exposures 15,000-32,000,000 times higher than typical environmental levels and cannot inform the shape of the dose-response at the environmental levels of interest (i.e., those being extrapolated to). While the TCEQ model itself can in theory be characterized as sublinear in nature, examination of the figures in the DSD (Appendix 5 of the revised DSD) show that the dose-response from the Cox proportional hazards model is actually indistinguishable from linear across doses of interest (i.e., the modeled dose-response is a straight line) and has no associated threshold. The Cox proportional hazards model is a standard dose-response model used throughout risk assessment (e.g., Steenland et al. papers, publications supporting USEPA NAAQS dose-response assessments, USEPA 2016).

Comment 12:

TCEQ's consideration of what it calls "endogenous exposure" levels of EtO creates in effect a threshold of exogenous exposure through inhalation of EtO. TCEQ disregards inhaled pollution and cancer risk resulting from that exposure at levels below that threshold. Specifically, TCEQ claims that EtO's effects "would be buffered by cellular repair mechanisms," "at doses near the endogenous range." TCEQ does not recognize that "any dose, no matter how small, increases the probability of causing an effect."

Response:

TCEQ's dose-response model is not a threshold model, but rather predicts excess risk all the way down to zero dose, recognizing that "any dose, no matter how small, increases the probability of causing an effect." The brief discussion of biological protective mechanisms in the DSD does not change that or affect TCEQ's rejection of an overall supra-linear model based on the absence of MOA data to adequately justify such a model. The discussion of endogenous EtO does demonstrate that relatively speaking, the internal exposures resulting from USEPA's risk-based values are orders of magnitude below normal endogenous doses and are therefore of implausible biological significance. To recognize this does not create a threshold in the TCEQ's dose-response assessment.

Comment 13:

TCEQ's model was considered by USEPA to have poor fit.

Response:

USEPA did not properly calculate model fit criteria (see Appendix 4 of the revised DSD). Briefly, USEPA (2016) did not account for statistically estimating the optimized knot value, making the degrees of freedom (*df*) inappropriately reduced for the spline models, which resulted in: (1) inappropriately decreased p-values for adequate statistical fit by spline models, incorrectly implying that the linear two-piece spline model for lymphoid cancer fit the data statistically better than other models; and (2) inappropriately decreased Akaike information criteria (AIC) for spline models, which did not allow for an appropriate comparison of model fit among models for lymphoid cancer (or breast cancer incidence). The knot values, being statistically estimated/optimized based on the NIOSH data, did not conform to the USEPA SAB's recommendation of potentially fixing some model parameters *not estimated from the data* in the interest of parsimony (see p. 12 of SAB 2015). The TCEQ presents corrected p-values for lymphoid cancer models (see Table 38 of the DSD), which indicate that the two-piece spline models do not explain the variability in the data statistically significantly better than the null model (zero slope) or the standard Cox regression model used by TCEQ. Additionally, the AIC values between most of the models are very similar (see Appendix 4 of the revised DSD).

Comment 14:

TCEQ has not cited and cannot point to any evidence demonstrating that the human body can negate cancer risk from inhaling any threshold level of EtO through a process of "cellular repair." In assuming endogenous levels are safe, TCEQ fails to consider that inhaling EtO results in exposure above and beyond any background endogenous exposure. The endogenous levels and background ambient concentrations of EtO likely contribute to the high background rates for both lymphatic and breast cancer incidence in the general population.

Response:

Please refer to TCEQ's response to Comment #12 above, which addresses the TCEQ's non-threshold dose-response model. In addition, the TCEQ does not consider endogenous levels as safe, just the opposite: the TCEQ quantifies risk to endogenous levels, and the dose-

response/URF based on the NIOSH cohort quantifies excess risk above and beyond background endogenous exposure. The TCEQ agrees that endogenous levels and background ambient air concentrations of EtO contribute to background EtO exposure for the general population, but TCEQ has also demonstrated that USEPA's dose-response model overestimates the general population background risk for lymphoid cancer and is therefore implausible, while results based on TCEQ's URF are plausible. The TCEQ assessment does not set a threshold or assume any amount of EtO, endogenous or exogenous, is without risk. In fact, it quantifies the risks from both.

Comment 15:

TCEQ relies on Kirman and Hays (2017) in regard to normal endogenous levels of EtO.

Response:

The TCEQ does use Kirman and Hayes (2017) to provide information about endogenous levels of EtO. These data are used to put the URF in context, and are not deterministic as to the rejection of USEPA's overall supra-linear dose-response model. The TCEQ will thoroughly consider any data or analyses that are provided that confirm, expand on, or rebut the peer-reviewed study of Kirman and Hays (2017). At the same time, the TCEQ notes that the reported mean human background endogenous HEV level of 21.1 pmol/g Hb appears reasonable given background HEV levels in control rats (≈42-50 pmol/g Hb) and mice (≈58-100 pmol/g Hb) (Walker et al. 1993, 2000).

Comment 16:

TCEQ acknowledges that children are more susceptible to the mutagenic effects of EtO, and claims to "include adjustments for [EPA (2005b) age-dependent adjustment factors (ADAFs)] using the approach described in Sielken and Valdez-Flores (2009)." However, USEPA found this study "misinterpreted the application of the [ADAFs] such that, even though they purported to apply the factors, this application had no impact on the risk estimate. TCEQ should apply a factor that is more protective than the 2016 IRIS value to account for the additional cancer risk a person faces over their lifetime due to cross-placental carcinogenic exposure during early fetal development. In 2009, California's Office of Environmental Health Hazard Assessment (OEHHA) published a review of the scientific literature surrounding prenatal susceptibility, and developed procedures for exposure assessment during fetal development. OEHHA specifically discusses the use of a 10X adjustment factor for cancer risk to account for prenatal (third trimester) to age 2 exposures. TCEQ should use the approach described by OEHHA to properly account for the effects of *in utero* exposure.

Response:

The TCEQ guidelines incorporate USEPA guidance on ADAFs, but not the referenced 10-fold OEHHA adjustment that USEPA ADAF guidance does not include. The TCEQ DSD now calculates the ADAF-adjusted ESL consistent with equation 5-17 of the TCEQ guidelines (TCEQ 2015), rounding it to two significant figures to produce an ESL of 2.4 ppb. This rounded value would be

insensitive to an additional 10-fold factor for the third trimester (i.e., with a 10-fold ADAF for *in utero* exposure, equation 5-17 would be 5.9E-06/URF = 5.9E-06/2.5E-06 per ppb = 2.36 ppb).

Comment 17:

TCEQ must consider breast cancer risk. TCEQ's proposed factor is invalid and unlawful because it fails to consider risks to women.

Response:

Please refer to the TCEQ's response to Comment #5 in regard to including breast cancer as an endpoint in this evaluation. The weight of evidence for EtO-induced breast cancer is now discussed in a new appendix to the revised DSD (Appendix 6).

In addition, the lymphoid cancer estimate is shown to be protective for women, because the lymphoid cancer effects in the NIOSH cohort (which included both male and female workers) were weaker in women than in men. The dose-response assessment evaluates lymphoid cancer risk to males and females combined as well as males alone. Analyses with both males and females combined compared to results based on males alone show that use of risk results based on males alone to predict lymphoid cancer in females generates a lower ESL and so is conservative. Thus, the long-term ESL and its associated URF used to evaluate lymphoid cancer risk in women will be protective for lymphoid cancer risk in women.

Comment 18:

TCEQ must consider cancer incidence, not only cancer mortality. TCEQ must consider a sufficient lifetime period of exposure of 85 years instead of 70 years. TCEQ must consider the disproportionate exposure of people of color and communities with multiple sources of EtO. TCEQ must follow all applicable civil rights law in this proceeding and all other actions. TCEQ has previously faced a complaint (No. 01R-00-R6) that it did not follow applicable requirements of Title VI of the Civil Rights Act, 42 U.S.C. § 2000d et seq., including public participation and calculations, in regard to an air permit modification. TCEQ entered into an agreement and under that agreement (see Section III), EPA is now requiring, and TCEQ has agreed, to hold at least two accessible community meetings to discuss opportunities for public involvement, with advance notice, accessibility, consideration of multilingual information and interpretation services for that process. TCEQ must ensure at least that it grants a public hearing and supports a similar level of public participation in this proceeding because TCEQ is proposing to use the ethylene oxide cancer risk factor for permitting proceedings. Failure to provide a meaningful opportunity for public participation here would contravene Title VI.

Response:

Partly in consideration of the issue of mortality versus incidence, the TCEQ utilized the 95% UCL on the slope, which is common in dose-response assessment. The TCEQ toxicity factor guidelines use the default exposure duration of 70 years, which is standard practice in risk assessment and consistent with TCEQ guidelines (TCEQ 2015). The long-term ESL applies equally for all people regardless of race, ethnicity, socioeconomic status, etc. and assumes 24

hours/day, 365 days/year exposure to be protective of even those most exposed to ambient EtO air concentrations.

The agreement referenced by the comment concerned a complaint from 2001 that specifically concerned changes at the ExxonMobil refinery in Beaumont, Texas. Details of the complaint concerned TCEQ public participation process in regard to the air quality permitting process. Specifically, USEPA accepted two parts of the complaint relating to the approval of a permit amendment without allowing the public an opportunity to participate in a contested case hearing, and approval of an air quality permit amendment with an increase in emissions in an area that already had problems with hydrogen sulfide and disparate adverse impact of a predominantly African American community. A global settlement between USEPA and TCEQ in July 2003 addressed the public participation allegation, detailing the changes to TCEQ's public participation requirements for permitting procedures. In 2017, TCEQ and USEPA resumed discussions, which resulted in the final agreement cited here by the commenter. That agreement required TCEQ to hold at least two public meetings in Beaumont to discuss both the public participation process for permitting and air quality monitoring issues in the area. These meetings were held on February 18, 2018 and April 26, 2018, and USEPA acknowledged on May 23, 2018 that TCEQ had met the requirements of the 2017 Informal Resolution.

Neither parts of the complaint accepted by USEPA have direct relevance to the current proposal, which is not a permitting action. The TCEQ is committed to protecting the health of the citizens of Texas and its environment. Discrimination on the basis of race, color, national origin, sex, or disability in the administration of TCEQ programs or activities is not allowed as required by federal and state laws and regulations. Although there are no TCEQ rules addressing environmental equity issues such as the location of permitted facilities in areas with minority and low-income populations, disparate exposures of pollutants to minority and low-income populations, or the disparate economic, environmental, and health effect on minority and low-income populations, the TCEQ has made a strong policy commitment to address environmental equity. The Office of the Chief Clerk works to help citizens and neighborhood groups participate in the regulatory process to ensure that agency programs that may affect human health or the environment operate without discrimination and to make sure that citizens' concerns are considered thoroughly and are handled in a way that is fair to all. Those interested may contact the Office of the Chief Clerk at 512-239-3300 for further information. More information on Environmental Equity may be found on the TCEQ website: https://www.tceq.texas.gov/agency/hearings/envequ.html.

Comment 19:

TCEQ must rely on scientifically sound, independent, peer-reviewed and published studies which the public can reasonably evaluate. TCEQ relied heavily on a dataset—the UCC cohort—for the derivation of unit risk estimates. TCEQ has thus far declined to share this study with the public.

Response:

The published UCC study was evaluated as part of TCEQ's thorough systematic review and analytical process. However, both the published UCC study and the currently unpublished update received zero weight in selection of the final URF, which was based exclusively on the NIOSH cohort. Please see the response to Comment #6 above in regard to the release of study information that is PII.

Comment 20:

IRIS follows the USEPA's Office of the Science Advisor's Principles of Scientific Integrity. These principles "ensure scientific integrity throughout the USEPA and promote scientific and ethical standards" throughout agency actions. Regarding conflicts of interest, "the Principles of Scientific Integrity sets forth the Agency's commitment to conducting science objectively, presenting results fairly and accurately, and avoiding conflicts of interest." USEPA's IRIS value follows these principles but thus far TCEQ's proposal does not.

Response

TCEQ maintains the highest scientific and ethical standards.

Comment 21:

TCEQ's proposed DSD must be independently, externally peer reviewed. TCEQ continues to withhold from the public the very studies and calculations it relies on and directly cites. At every step, TCEQ has sided with industry.

Response

The TCEQ is conducting an external peer review for our draft EtO DSD in the first quarter of 2020. All studies relied on for the long-term ESL/URF are available to the public via the peer-reviewed literature, except for the update of the UCC cohort, which as noted above was not used to derive the URF. All methods and associated results are reported in the DSD. Lastly and importantly, the TCEQ sides with the best science available, regardless of who may stand with the agency in favor of the best science in any given circumstance. The agency follows wherever the weight of scientific evidence leads.

Comments from Dr. Kyle Steenland

Dr. Steenland submitted several comments (see below). Dr. Steenland, along with Dr. Stayner and other NIOSH colleagues, conducted the original epidemiologic studies providing the key data investigating the association of EtO with lymphoid and breast cancer and has conducted several risk assessments for EtO and cancer, some under contract with USEPA in support of their 2016 risk assessment.

Comment 1:

A choice not to include risks to women from breast cancer in the quantitative assessment is an error. NIOSH conducted a breast cancer incidence study which clearly showed a significant positive exposure response for EtO, using an appropriate lag. Several smaller studies have supported this finding. Incidence is preferable to mortality for many outcomes, including breast cancer and hematopoetic cancer. TCEQ is choosing to simply ignore the breast cancer findings. I note that breast cancer incidence data also support a supra-linear exposure-response model, and that the incidence data are supported by breast cancer mortality findings.

Response:

The TCEQ evaluated breast cancer as a candidate cancer endpoint. USEPA's assessment, however, is driven by lymphoid cancer as the primary contributor to the URF. Subsequent to USEPA's 2016 assessment and TCEQ's systematic review of the peer-reviewed literature for EtO, recent meta-analyses of available studies have been published (Marsh et al. 2019, Vincent et al. 2019). These new meta-analyses included Steenland et al. (2003, 2004) and the smaller Mikoczy et al. (2011) study cited by USEPA (2016), as well as other studies, and reported breast cancer meta-RRs of 0.97 (0.80, 1.18) (Marsh et al. 2019) and 0.92 (0.84, 1.02) (Vincent et al. 2019). The Marsh et al. study concluded [*emphasis added*], "Evaluations of workers exposed during sterilization processes *do not support the conclusion that EO exposure is associated with an increased risk of breast cancer*." Similarly, the Vincent et al. (2019) study concluded, "Higher quality epidemiological studies demonstrated no increased risk of breast cancers." In addition to evaluating epidemiological evidence, Vincent et al. (2019) also evaluated animal study results and concluded that they provide no strong indication that EtO causes mammary tumors. These recent meta-analyses and other information (IARC 2019) further support TCEQ's decision not to base the URF on breast cancer. Human data are by far the most relevant for derivation of human toxicity factors, and the human data themselves are inconclusive (as acknowledged by USEPA 2016). The weight of evidence for EtO-induced breast cancer is now discussed in a new appendix (Appendix 6) to the revised DSD.

Simply for additional perspective, the cumulative exposure associated with a lifetime of environmental exposure to the revised long-term ESL of 2.4 ppb (≈60 ppm × days) is around 250 times less than the lowest cumulative exposure levels associated with a statistical increase in breast cancer incidence in the NIOSH cohort (>14,620 ppm-days, 15-year lagged exposure; see Table 6 of the revised DSD). This statistical increase in breast cancer incidence only occurred in the highest (5th) exposure group. If the underlying dose-response for EtO-induced breast cancer in humans were supra-linear, statistically significant increases in critical cancer

endpoints would be expected beginning in the lower occupational exposure groups. That is, if exogenous EtO had a steep dose-response slope (i.e., were a potent carcinogen) in the true low-dose region, such as near the range of endogenous levels, then statistically increased breast cancer rates would be expected at the "low" worker doses evaluated for the large NIOSH cohort, particularly considering that even "low" historical worker exposures have been significant.

Comment 2:

A choice to use the linear model among a variety of models for lymphoid cancer mortality that were previously examined by USEPA, even though this model did not fit as well as the other models which show that the exposure-response is supra-linear, is a major error (e.g., the two piece spline and the log cumulative dose model, both of which showed good fit to the data, $p<0.05$, unlike the linear model, see Appendix D in the USEPA's risk assessment). The linear model results in much lower estimates of risk in the low dose region, which is the region of interest.

Response:

The TCEQ disagrees that use of "the linear model" for lymphoid cancer mortality is an error. The commenter indicates that the reason for this is that the supra-linear two-piece spline model and log cumulative model show good fit to the data whereas the linear model does not, citing USEPA (2016). However, the USEPA did not properly calculate model fit criteria (see Appendix 4 of the revised DSD). Briefly, USEPA (2016) did not account for statistically estimating the optimized knot value, making the degrees of freedom (*df*) inappropriately reduced for the spline models, which resulted in: (1) inappropriately decreased p-values for adequate statistical fit by spline models, incorrectly implying that the linear two-piece spline model for lymphoid cancer fit the data statistically better than other models; and (2) inappropriately decreased Akaike information criteria (AIC) for spline models, which did not allow for an appropriate comparison of model fit among models for lymphoid cancer (or breast cancer incidence). The knot values, being statistically estimated/optimized based on the NIOSH data, clearly did not conform to the USEPA SAB's recommendation of potentially fixing some model parameters *not estimated from the data* in the interest of parsimony (see p. 12 of SAB 2015). The TCEQ presents corrected p-values for lymphoid cancer models (see Table 38 of the DSD), which indicate that the two-piece spline models do not explain the variability in the data statistically significantly better than the null model (zero slope) or the standard Cox regression model used by TCEQ. Additionally, the AIC values between most of the models are very similar (see Appendix 4 of the revised DSD). In addition, because the MOA data do not scientifically justify use of a supra-linear model (i.e., the steep lower-dose slope) for low-dose extrapolation (see Section 3.4.1.4.1 of the DSD), the two-piece spline models were not considered by the TCEQ for adoption; nor were other models that have an inherently supra-linear dose-response over the exposure range (i.e., log-linear or linear models with log cumulative exposure or with square-root transformation of cumulative exposure).

Comment 3:

The TCEQ interpretation of data on protein adduct levels associated with EtO exposure is misleading. We don't know if endogenous (internal) levels of EtO contribute to background levels of cancer, but we do know that increasing them with exogenous (external) exposures leads to increased cancer. That is what is important. Furthermore, the animal (positive rat and mice studies) and mechanistic data (mutagenicity, effect on chromosomal aberrations) all support the positive human data. That combined evidence is what IARC has determined that EtO is a definite human carcinogen.

Response:

Consistent with the comment, the TCEQ's DSD adopts USEPA's/IARC's carcinogenic classification and treats EtO as a mutagen that is *carcinogenic to humans*. The TCEQ uses the data on protein adduct levels associated with EtO exposure to provide valuable context, although it is not deterministic in terms of the dose-response model selected. The endogenous EtO protein adduct data are a secondary consideration in the overall weight of evidence, with the primary consideration for TCEQ's selection of a standard dose-response model being driven by the mechanistic data that shows a standard mutagenic MOA that is not expected to cause a supra-linear dose-response. Even in the absence of data on EtO adduct background levels normally found in the population, TCEQ's model selection would not change. For example, since lymphoid cancer drove the TCEQ and the USEPA carcinogenic assessment, perhaps the most relevant mutagenicity data was that in the bone marrow of mice exposed to EtO by inhalation *in vivo* (Recio et al. 2004), which USEPA indicates is consistent with a linear dose-response (see C-17 of USEPA 2016).

The TCEQ's use of background/endogenous EtO adduct data is consistent with the USEPA using their URF to extrapolate to background endogenous doses and well below, and to USEPA's acknowledgement that endogenous EtO may contribute to background risk (pp. 4-95 to 4-96 of USEPA 2016), which makes sense for an endogenously-produced mutagen. Furthermore, in dose-response/risk assessment it is standard practice to consider equal internal doses as equipotent in producing both noncarcinogenic and carcinogenic effects (i.e., that equal internal doses of a chemical produce equivalent risk). This is scientifically reasonable, which is why it is standard practice, absent a scientifically robust demonstration that this standard assumption should not apply (e.g., mechanistic or toxicokinetic information). Thus, the standard risk assessment practice of considering equal internal doses as equipotent in producing carcinogenic effects underlies TCEQ's application of the USEPA and TCEQ URFs to internal endogenous doses of the same mutagenic chemical.

Regarding animal data, while laboratory animal data can help provide support, interspecies differences in carcinogenic responses such as tumor types and sensitivity are common, even between different rodent species (e.g., EtO-induced mammary tumors in mice but not rats). For example, IARC (2019) analyzed tumor site concordance using a dataset of the 111 distinct Group 1 (carcinogenic to humans) agents. Sixty agents had both a human tumor site and an animal tumor site identified and were used to evaluate concordance across 39 tumor sites in animals and humans (see Figures 21.1 and 21.2 of IARC 2019). Reported results show that

breast cancer, for example, is more frequently/commonly induced in laboratory animal species by these agents than in humans. More telling is that while there is 47% overlap between agents that cause lymphoid and hematopoietic cancers in humans and animals, there is only 20% overlap between agents that have been shown to cause breast cancer in humans and animals (Table 21.7 of IARC 2019). The IARC (2019) unanimous consensus statement was that "*At present, the state of the science does not support tumour site concordance as a general principle*."

Accordingly, animal data are not deterministic as to the sites of chemically-attributable carcinogenesis in humans; even more so when laboratory animal results are inconsistent (e.g., mammary tumors in mice but not rats) and the human database is relatively robust. For example, lung cancer was statistically increased in both male and female EtO-exposed mice at incidences of 53% and 45%, respectively (Table 3-3 in USEPA 2016), but is not a candidate endpoint in humans as data for this very strong carcinogenic response in mice is simply not predictive for humans (i.e., no interspecies site concordance; SMR of 1.05 (0.95, 1.17) in Table 1 of Steenland et al. 2004). Similarly, EtO statistically significantly increased brain tumors in rats of both sexes (Table 3-5 in USEPA 2016), but yet again, these results are not predictive for humans. In fact, brain cancer for the NIOSH cohort is statistically significantly decreased (i.e., SMR of 0.59 (0.36, 0.91) in Table 1 of Steenland et al. 2004), just the opposite of what the rat data would suggest. *Clearly, laboratory animal data for EtO-induced cancers cannot be relied upon to identity cancer sites in humans.*

Comment 4:

Extrapolating curve-fit model down to environmental dose. The Texas draft in contrast to the IRIS EtO assessment relies on direct extrapolation of the selected models from high to down to low (i.e., environmental) dose. This approach contrasts with the emphasis in USEPA's Cancer Guidelines on limiting use of curve fit dose response models to the observable range – to support estimation point of departure (BMDL) that can then be used for straight line extrapolation to low dose when appropriate (as for a direct acting mutagenic carcinogen like EtO). Choice of the better fitting model (the supra-linear model), then extrapolating down from a point of departure, is the best approach to EtO risk assessment.

Response:

TCEQ evaluated the lowest effective concentration (LEC) at an extra risk of 1 in a 100,000 consistent with USEPA cancer guidelines (2005) on the selection of a POD at the low-end of the observable range of exposures. Although for animal studies a typical POD is an extra risk of 0.10 because it corresponds to doses near the low-end of the doses, in epidemiological studies a lower level of risk needs to be used.

TCEQ used the standard Cox proportional hazards model to calculate the LEC for an extra risk of 1 in a 100,000 because the effective concentrations (ECs) corresponding to the same risk level are in the range of the observed data in the NIOSH study. That is, the EC for an extra risk of 1 in 100,000 of lymphoid cancer mortality in males is 9.67E-03 ppm for 70 years with an exposure lag of 15 years, which correspond to a cumulative

occupational exposure of 591 ppm-days. There are 7 male workers in the NIOSH cohort with cumulative exposures less than 591 ppm-days. That is, 25.9% of the male workers in the NIOSH cohort that died with lymphoid cancer were exposed to cumulative exposures of less than the EC for a 1 in a 100,000 excess risk. In contrast, the EC for 1 in 100 results in environmental concentrations corresponding to cumulative occupational exposures of 354,400 ppm-days, which exceeds the largest cumulative exposure of lymphoid male decedents in the NIOSH study.

Table 1 in TCEQ's response to ACC Comment #15 (above) shows the EC corresponding to different risk levels and the corresponding cumulative exposures with the number of lymphoid mortality cases of the male workers in the NIOSH study. The results in Table 1 show that the EC for an extra risk of 1 in a 100 is outside the range of cumulative exposures for the male lymphoid mortalities observed in the NIOSH study and in the upper 1% of cumulative exposures for all male workers. That is, all males that died with lymphoid cancers and more than 99% of all male workers had cumulative exposures less than EC (1/100). Thus, the NIOSH study does not support an extra risk of 1 in a 100 as a point of departure.

The EC for an extra risk of 1 in a 1,000 is a concentration that is in the high-end of cumulative exposures of male lymphoid mortalities observed in the NIOSH study. That is, 77.78% of all males that died with lymphoid cancers and 94.48% of all male workers had cumulative exposures less than the EC (1/1,000). Thus, a point of departure of 1 in 1,000 is at the higher-end of the cumulative exposures of male workers of the NIOSH study. The EC for an extra risk of 1 in 10,000 is a concentration that includes 48.15% of the decedent men with lymphoid cancer and 66.45% of all men in the NIOSH cohort with smaller cumulative exposures. The EC for an extra risk of 1 in 100,000 includes 25.93% of male lymphoid decedents and 30.17% of all males in the NIOSH study with smaller cumulative exposures. Thus, use of an extra risk of 1 in 100,000 is supported by the NIOSH observed data, being near the lower end of the observed range of cumulative exposures to EtO, and is consistent with USEPA and TCEQ guidelines (USEPA 2005a, TCEQ 2015) on the selection of a POD at the low-end of the observable range of exposures.

Based on Table 1 results, using either 1 in 10,000 or 1 in 100,000 extra risk PODs (as PODs in the range of the observed data and close to the low-end of the observable range) round to the same ADAF-unadjusted URF selected by the TCEQ (2.5E-06 per ppb). Looking at it from a different perspective, using the 1 in 10,000 excess risk LEC of 4.04E-02 ppm as the POD and linear extrapolation, the 1 in 100,000 air concentration (ADAF unadjusted) is still 4 ppb (i.e., 1E-05/2.47E-06 per ppb = 4.05 ppb).
This information has been added to Appendix 7 of the revised DSD.

Comment 5:

The draft's approach to comparing of observed and expected rates in the NIOSH cohort is incorrect. The use of national tumor rates to predict cancers in the NIOSH worker cohort is

inappropriate because it ignores the healthy worker effect, whereby working populations have lower mortality rates than the national population. The use of internal comparisons, as used by NIOSH investigators and USEPA risk assessors, avoids this issue, and is standard in occupational risk assessment. The models used by USEPA predict quite well the observed occurrence of cancer in the cohort.

Response:

The approach followed by TCEQ to compare the observed and expected number of deaths is correct, as it is the same approach used to calculate standard mortality ratios (SMRs). The approach for calculating SMRs is well established and has been used by regulatory agencies and researchers to compare mortality rates in target populations to mortality rates in reference populations. Thus, the approach used is well documented and has been extensively used.

Addressing the use of internal comparisons, the models used by TCEQ were also derived using internal comparisons and did not rely on the general U.S. population standard mortality rates. As indicated by Dr. Steenland, this is standard in exposure response models fit to epidemiological studies based on occupational cohorts.

To Dr. Steenland's assertion about the USEPA's model's predicting the observed occurrence of cancer in the cohort, the TCEQ cannot confirm this statement for two reasons. First, the USEPA's risk assessment document does not report any comparison between the number of cases in the NIOSH cohort and what is predicted by their models. Secondly, importantly the TCEQ did conduct the relevant analysis and found that if USEPA's chosen model were correct, then there should have been statistically significantly more lymphoid cancer mortalities in the NIOSH studies than the 53 that were actually observed, with statistically significant increases in every exposure quintile that in fact did not occur.

Lastly, we address the potential of the healthy worker effect to significantly affect TCEQ's calculations of model predictions (i.e., two-piece linear spline, Cox proportional hazards model) for the underlying lymphoid cancers in the NIOSH cohort (in Figures 8-12 and Appendix 2 of the revised DSD). Though opinions vary about using general population background rates for evaluating cause-specific mortality rates of occupational studies, it is standard practice/methodology in epidemiology studies to use general population background rates to evaluate the well-established and widely-accepted SMRs and SIRs because there is often no scientific evaluation of the magnitude of the healthy worker effect. In general, the healthy worker effect, if any, is cause-specific and cannot be ascertained very easily. However, Kirkeleit et al. (2013) researched the healthy worker effect in a large study of 366,114 randomly selected Norwegian workers and compared the incidence of numerous endpoints with the general population. Their findings indicate that there is potential for a healthy worker effect for some endpoints and increased incidence ("unhealthy" worker effect) for other endpoints. Kirkeleit et al. (2013) indicates that the healthy worker effect for mortality from cancer is minimal, if any, overall. Relevant to the EtO assessment, for lymphoid and hematopoietic cancer incidence Kirkeleit et al. (2013) did not find a healthy worker effect, with SIRs and 95% confidence intervals of 0.97 (0.90, 1.03) and 1.09 (0.92, 1.27) for male and female workers, respectively.

The lack of a health worker effect was also true for breast cancer with an SIR and 95% confidence interval of 1.02 (0.95, 1.09).

The SMRs for the unexposed individuals in the NIOSH study indicate that the lymphoid mortality rate in unexposed male and female workers in the NIOSH study are not statistically significantly different from the lymphoid mortality rate of the general U.S. population. A footnote to Table 32 in the proposed TCEQ DSD states "Quintile 1 is the control (unexposed lagged-out) group with 9 lymphoid cancer mortalities observed and 11.5 mortalities predicted by all models with a 95% confidence interval of (6.0 and 25.2), which includes the 9 lymphoid cancer deaths." This confidence interval indicates that the observed 9 lymphoid cancer deaths in the unexposed male and female workers of the NIOSH cohort is consistent with the expected number of lymphoid cancer deaths in the general U.S. population during the same period of time after accounting for age, sex, and calendar year. Expressed in terms of SMRs, the SMR for lymphoid cancer deaths in the unexposed male and female NIOSH workers is equal to 0.78 (9/11.5) with a 95% confidence interval (CI) equal to (0.36, 1.50). The 95% CI on the SMR for unexposed workers includes the value of one, which indicates that the mortality rate in the unexposed workers in the NIOSH study and the U.S. population mortality rate are not statistically significantly different at the 5% significance level (p-value of 0.29). Similar results are obtained for the male NIOSH workers that drive lymphoid cancer risk: the SMR for lymphoid cancer deaths in the unexposed male NIOSH workers is equal to 1.03 (6/5.8) with a 95% CI of (0.38, 2.25). Thus, the lymphoid cancer mortality rate in unexposed male workers in the NIOSH cohort, the gender that drives the URF, is not statistically significantly different than that in the U.S. population (p-value of 0.64). In summary, these results demonstrate that there is no healthy worker effect for this critical endpoint in this key group (i.e., male workers, who drive lymphoid cancer risk in the cohort) or in males and female workers combined. These results based on the NIOSH cohort are consistent with the findings of Kirkeleit et al. (2013).

Despite the discussion above about the likely lack of a healthy worker effect for the endpoints of interest in the NIOSH study, to be thorough the TCEQ conducted a sensitivity analysis *assuming a healthy worker effect* for cancer mortality. Kirkeleit et al. (2013) estimates an overall cancer SMR of 0.85 and 0.84 for male and female workers, respectively. For purposes of a sensitivity analysis, the TCEQ assumed that these overall cancer SMRs apply to lymphoid cancers. That is, despite data to the contrary, the TCEQ sensitivity analysis assumes NIOSH workers were "healthier" than the general population as to cancer mortality by multiplying the U.S. male and female background hazard rates by 0.85 and 0.84, respectively, to account for the assumed healthy worker effect. *The results did not change significantly.* The maximum likelihood estimate (MLE) of USEPA's selected model still statistically significantly overestimates the number of observed (53) lymphoid deaths in the NIOSH study, predicting 77.5 lymphoid deaths with a 95% CI of (59.3, 103.6). By contrast, the number of lymphoid deaths in the NIOSH study (53) is encompassed by the estimate from the standard Cox proportional hazards model used by the TCEQ: 44.3 with a 95% CI of (33.9, 59.2). Thus, even conservatively assuming a healthy worker effect (in the face of more study-specific data to the contrary), USEPA's selected model significantly overestimates the observed data.

Comment 6:

The use of the UCC cohort as equivalent in importance to the NIOSH cohort is inappropriate. The UCC cohort was much smaller (2000 vs 18,000) and had far less developed exposure estimates, making conclusions about exposure-response in that cohort less valuable. This is why the EPA and its Scientific Advisory Board (SAB) recommended reliance on the NIOSH cohort.

Response:

The TCEQ agrees with Dr. Steenland and the SAB that the NIOSH cohort results should receive much more weight than the UCC cohort. In selecting the final URF the exposure-response analyses for the UCC cohort received zero weight in the TCEQ analysis (partially due to the cohort's smaller size). Thus, consistent with the SAB recommendation, the TCEQ URF exclusively relies on the NIOSH study.

Comments from Miscellaneous Emails, Letters, and Postcards

The TCEQ received numerous emails and letters regarding the proposed DSD for EtO. While the correspondence did not contain specific scientific comments, common comments or themes are addressed by the TCEQ below.

Comment 1:

You suggest that the EPA over-estimated the cancer risk factor in its original assessment of EtO. Stop downplaying toxic cancer risks of EtO.

Response:

The TCEQ's DSD statistically demonstrates that our model relatively accurately predicts risk while USEPA's selected model assessment statistically significantly overestimates risk.

Comment 2:

Fully consider the health of Texas children and vulnerable communities and retract your dangerous proposal to Environmental Protection Agency to lower the national risk factor of EtO. It is better to err on the side of safety.

Response:

Consistent with USEPA guidance, the TCEQ uses ADAFs to account for the potentially increased susceptibility of children. Additionally, the TCEQ uses the 95% UCL to err on the side of safety and help ensure that vulnerable people and communities are protected against the potential carcinogenic effects of EtO in ambient air. Lastly, the TCEQ's DSD documents the derivation of a URF for use in air permits in Texas and is not a proposal directed to the USEPA.

Comment 3:

The report is flawed and incomplete in its risk assessment because it uses a male-only occupational study on lymphoid cancer to support its proposed risk value and does not account for the risk of breast cancer and other serious health risks EtO exposure poses (e.g., potential effects on human fetuses and child development).

Response:

The TCEQ DSD is restricted to a carcinogenic dose-response assessment, similar to how the USEPA (2016) assessment is limited to carcinogenic risk. Protecting against cancer endpoints is expected to also be protective of other health effects such as developmental effects.

The TCEQ URF is ultimately based on the same key epidemiological study as USEPA's URF, which included both male and female workers, not male workers only. The DSD evaluates lymphoid cancer risk to males and females combined as well as males alone, and the results demonstrate that to include female results in the final URF would actually make it less conservative for both males and females (i.e., the long-term ESL would be higher). Analyses with both males and females combined compared to results based on males alone show that

use of risk results based on males alone is conservative for predicting risk in females. While technically less accurate, the TCEQ chose to be more conservative and use the male-only analysis because using male and female results combined would result in a less conservative (i.e., higher) long-term ESL value.

Lastly, the TCEQ evaluated breast cancer as a candidate cancer endpoint. USEPA's assessment, however, is driven by lymphoid cancer as the primary contributor to the URF. Subsequent to USEPA's 2016 assessment and TCEQ's systematic review of the peer-reviewed literature for EtO, recent meta-analyses of available studies have been published (Marsh et al. 2019, Vincent et al. 2019). These new meta-analyses included Steenland et al. (2003, 2004) and the smaller Mikoczy et al. (2011) study cited by USEPA (2016), as well as other studies, and reported breast cancer meta-RRs of 0.97 (0.80, 1.18) (Marsh et al. 2019) and 0.92 (0.84, 1.02) (Vincent et al. 2019). The Marsh et al. study concluded [*emphasis added*], "Evaluations of workers exposed during sterilization processes *do not support the conclusion that EO exposure is associated with an increased risk of breast cancer*." Similarly, the Vincent et al. (2019) study concluded, "Higher quality epidemiological studies demonstrated no increased risk of breast cancers." In addition to evaluating epidemiological evidence, Vincent et al. (2019) also evaluated animal study results and concluded that they provide no strong indication that EtO causes mammary tumors. These recent meta-analyses and other information (IARC 2019) further support TCEQ's decision not to base the URF on breast cancer. Human data are by far the most relevant for derivation of human toxicity factors, and the human data themselves are inconclusive (as acknowledged by USEPA 2016). The weight of evidence for EtO-induced breast cancer is now discussed in a new appendix to the revised DSD (Appendix 6).

Comment 4:

EtO is a toxic mutagen and known potent human carcinogen. I don't want this in my community or anyone else's in the USA. Environmental standards should be kept at levels that are far below a dangerous level.

Response:

The TCEQ's assessment treats EtO as a potent toxic mutagen that is *carcinogenic to humans* (i.e., a known human carcinogen). For example, TCEQ's revised long-term ESL for EtO (2.4 ppb, ADAF adjusted) puts it on the same order of magnitude level as benzene (1.4 ppb), another relatively potent known human carcinogen.

Typical EtO ambient air concentrations around the US represent an insignificant carcinogenic risk, especially considering the EtO doses normally produced endogenously in all humans. The TCEQ strongly agrees that environmental regulatory levels should be far below dangerous levels, and therefore TCEQ's revised 2.4 ppb value: (1) corresponds to an internal dose within the range of EtO doses normal produced endogenously in the human body; and (2) corresponds to a lifetime cumulative exposure (61 ppm × days) that is orders of magnitude below air concentrations found in the key epidemiological study (used by TCEQ/USEPA) to be associated with either increased lymphoid or breast cancer risk (see Table 5 and Figure 3 of the DSD).

Comment 5:

TCEQ's flawed proposal ignores the best available science on toxicology. The TCEQ Toxicology Division needs to acknowledge the flaws in its internal EtO assessment and recognize the serious dangers that its weakened EtO risk factor would have on the state. The EPA has already completed a rigorous 10-year review of EtO potency through the Integrated Risk Information System (IRIS) program that has a more substantial scientific basis than the TCEQ's proposal for determining EtO's risk factor. I support the findings and conclusions of the IRIS chemical assessment.

Response:

The TCEQ's DSD fully considers the USEPA's derivation of their EtO URF and finds several serious flaws in USEPA's 2016 assessment. The DSD then carefully and completely lays out the scientific basis for our EtO URF derivation, based on the best available science and supported by multiple lines of evidence, some of which are newly available. The revised DSD and accompanying analyses will undergo peer review in the first quarter of 2020.

Comments from the University of California, San Francisco

The authors of these comments collectively declared no direct or indirect financial or fiduciary interest in any chemical or product that is the subject of these comments. The TCEQ notes that the first author of these comments is also the first author on USEPA (2016).

Comment 1:

The DSD's final risk estimate does not include breast cancer risks.

EPA's conclusion of a potential breast cancer hazard from exposure to ethylene oxide (EtO) was supported by the SAB. TCEQ seems to acknowledge a potential breast cancer hazard and considers EPA's quantitative risk estimates for breast cancer, but then rejects EPA's estimates and includes no alternative estimates for breast cancer.

The SAB explicitly endorsed EPA's use of a two-piece spline model for modeling the breast cancer incidence data, and EPA's unit risk estimates for breast cancer incidence are based on this model. TCEQ's rationales for rejecting EPA's approach are flawed because TCEQ conflates endogenous (background) exposures with low exogenous exposures, assuming that small increases in exposure above background would not be biologically meaningful, despite the fact that breast cancer has relatively high background rates.

There is uncertainty about the risks at low levels of exposure, and this is why EPA applies a linear extrapolation from models derived in the observable range of the data. Use of linear low-exposure extrapolation was supported by the established mutagenic mode of action (MOA) and the SAB. These issues are discussed in more detail below (see Comment #3, 4, 6, 11e).

Having rejected EPA's human-based breast cancer risk estimates in the proposed DSD, TCEQ could have considered the rodent-based estimates presented by EPA, rather than completely discounting breast cancer risk. Indeed, in March 2017 TCEQ did exactly that, adopting a value of $7.6x10^{-5}$ per $\mu g/m^3$, the EPA IRIS value for total cancer risk based on rodent data (see Table and Appendix A). Yet, TCEQ's 2019 total risk estimates are for lymphoid cancers only, and the DSD does not provide a valid scientific rationale for not including breast cancer risks in the final unit risk estimate. Because the DSD fails to include breast cancer, TCEQ's final risk estimate is a major underestimation of the actual cancer risks posed by ethylene oxide.

Response:

The TCEQ evaluated breast cancer as a candidate cancer endpoint. USEPA's assessment, however, is driven by lymphoid cancer as the primary contributor to the URF. Subsequent to USEPA's 2016 assessment and TCEQ's systematic review of the peer-reviewed literature for EtO, recent meta-analyses of available studies have been published (Marsh et al. 2019, Vincent et al. 2019). These new meta-analyses included Steenland et al. (2003, 2004) and the smaller Mikoczy et al. (2011) study cited by USEPA (2016), as well as other studies, and reported breast cancer meta-RRs of 0.97 (0.80, 1.18) (Marsh et al. 2019) and 0.92 (0.84, 1.02) (Vincent et al. 2019). The Marsh et al. study concluded [*emphasis added*], "Evaluations of workers exposed during sterilization processes *do not support the conclusion that EO exposure is associated with*

an increased risk of breast cancer." Similarly, the Vincent et al. (2019) study concluded, "Higher quality epidemiological studies demonstrated no increased risk of breast cancers." In addition to evaluating epidemiological evidence, Vincent et al. (2019) also evaluated animal study results and concluded that they provide no strong indication that EtO causes mammary tumors. These recent meta-analyses and other information (IARC 2019) further support TCEQ's decision not to base the URF on breast cancer. Human data are by far the most relevant for derivation of human toxicity factors, and the human data themselves are inconclusive (as acknowledged by USEPA 2016). The weight of evidence for EtO-induced breast cancer (i.e., "valid scientific rationale for not including breast cancer") is now discussed in a new appendix (Appendix 6) to the revised DSD.

USEPA (2016) acknowledges that human data are insufficient to establish that EtO is a human breast cancer carcinogen, which would be quite unexpected if EtO were in fact as highly potent of a carcinogen as USEPA (2016) purports given the large group of workers (including women) exposed to very high concentrations of EtO on a daily basis. As a result, USEPA must rely on support by laboratory animal studies in classifying EtO as carcinogenic to humans. However, upon closer inspection, the sites of EtO-induced cancers in animal models are of questionable direct human relevance for being predictive of, and therefore being used as confirming evidence for, the site(s) of human cancers. See the response to Comment #3 from Dr. Steenland above for EtO-specific examples of the lack of site concordance between tumors induced by EtO at high incidence in laboratory animals and epidemiological findings (i.e., lung cancer, brain cancer). Interspecies differences in carcinogenic responses such as tumor types and sensitivity are common, even between different rodent species (e.g., EtO-induced mammary tumors in mice but not rats). IARC (2019) analyzed tumor site concordance using a dataset of the 111 distinct Group 1 (carcinogenic to humans) agents. Sixty agents had both a human tumor site and an animal tumor site identified and were used to evaluate concordance across 39 tumor sites in animals and humans (see Figures 21.1 and 21.2 of IARC 2019). Reported results show that breast cancer, for example, is more frequently/commonly induced in laboratory animal species by these agents than in humans. More telling is that while there is 47% overlap between agents that cause lymphoid and hematopoietic cancers in humans and animals, there is only 20% overlap between agents that have been shown to cause breast cancer in humans and animals (Table 21.7 of IARC 2019). The IARC (2019) unanimous consensus statement was that "*At present, the state of the science does not support tumour site concordance as a general principle*." As a result, a URF based on mammary tumors in laboratory animals as a surrogate for human breast cancer, for which data are themselves inadequate to establish EtO as a human carcinogen (USEPA 2016), would be considered highly uncertain. [The 1.5-page March 2017 TCEQ EtO document referred to was merely an interim step prior to TCEQ being able to conduct a full review and produce a thorough DSD based on much more detailed analyses.]

Based on this information, it cannot be said that TCEQ's final risk estimate is an underestimation of the actual cancer risks posed by EtO, since it is not at all clear that EtO causes human breast cancer. In fact, the overall weight of evidence is against it (see Appendix 6 to the revised DSD). USEPA's *carcinogenic to humans* classification is best supported by the

lymphoid cancer data (e.g., NIOSH cohort), and TCEQ's final URF is best based on lymphoid cancer as the critical cancer endpoint.

Regarding endogenous EtO levels, assuming EtO does increase breast cancer risk despite the overall weight of evidence (see Appendix 6 to the revised DSD): given the background rate for breast cancer, exogenous EtO exposure resulting in higher than background internal doses would be more likely to be significant if endogenous EtO were known to be *the* cause of relatively high breast cancer background rates, but this is not the case. However, any risk associated with exogenous EtO exposures resulting in internal doses well within the range of normal endogenous background, particularly orders of magnitude below the 1st percentile of endogenous, will be indistinguishable from background biologically and inconsistent with the concept of excess risk. The TCEQ uses the data on protein adduct levels associated with EtO exposure to provide valuable context; it is not deterministic in terms of the dose-response model selected. The endogenous EtO protein adduct data are a secondary consideration in the overall weight of evidence, with the primary consideration for TCEQ's selection of a standard dose-response model being driven by the mechanistic data that shows a standard mutagenic MOA that is not expected to cause a supra-linear dose-response. Even in the absence of data on EtO adduct background levels normally found in the population, TCEQ's model selection would not change. For example, since lymphoid cancer drove the TCEQ and the USEPA carcinogenic assessment, perhaps the most relevant mutagenicity data was that in the bone marrow of mice exposed to EtO by inhalation *in vivo* (Recio et al. 2004), which USEPA indicates is consistent with a linear dose-response (see C-17 of USEPA 2016).

Lastly, consistent with the comment, the TCEQ uses linear low-dose extrapolation based on a mutagenic MOA (the Cox model is indistinguishable from linear over the doses of interest, as shown in the DSD, and a modeled POD is then used for URF derivation/linear extrapolation).

Comment 2:

The DSD discounts the role of expert peer review.

EPA's EtO carcinogenicity assessment was the subject of extensive review. In addition to review by other offices in EPA and other agencies in the federal government, the assessment twice underwent external peer review by EPA's independent SAB, which included discussions at open public meetings in 2006 and 2014; the SAB also considered public comments made at the meetings. In addition to addressing the SAB's comments, EPA considered public comments made at the 2006 and 2014 SAB meetings as well as at a public meeting in 2013. For the 2014 review by the SAB, the Board set up a panel of 14 experts from a range of relevant disciplines. After the review, the panel's report was reviewed and endorsed by the larger SAB.

As described in the comments below, the SAB explicitly endorsed EPA's approaches and rejected the model ultimately chosen by TCEQ, where the Commission's conclusions and approaches differed from those of EPA (e.g., discounting the breast cancer models, rejecting two-piece spline models, not using linear low-exposure extrapolation). The DSD does not present new data or evidence that was not considered by the SAB, nor does it provide an

appropriate scientific explanation for the significant departures from EPA's methodology. In contrast to the Agency, academic and public expert input and extensive peer review of the EPA assessment, the DSD has not been peer reviewed or subject to any external comments.

Response:

The TCEQ does not discount the value of external expert peer review. The TCEQ DSD has undergone internal peer review, and in addition to soliciting public comments and addressing to those public comments, the TCEQ assessment will undergo an external peer review by a panel of experts in relevant fields of study in the first quarter of 2020.

USEPA's SAB had limited information and was not provided all the data needed to validate/invalidate USEPA's dose-response models. Appendix 4 of TCEQ's assessment develops a corrected analysis of model fit criteria (i.e., p-values and AIC values). This analysis showed that once USEPA's omission of a degree of freedom in the statistical evaluation of their two-piece spline models is corrected, the two-piece spline model does not fit the data statistically better than the standard Cox proportional hazards model (or even the null model with zero slope).

TCEQ also developed other new analyses not available to the SAB. Appendix 2 of the TCEQ revised DSD compares the number of lymphoid cancer deaths observed in the NIOSH epidemiological study with the number of lymphoid cancer deaths that would be expected in the NIOSH epidemiological study if the exposure-response models were correct. A good model should be able to estimate the data that were used to fit the model; that is, for example, the number of lymphoid deaths observed in the NIOSH epidemiological study. However, USEPA's selected model assessment statistically significantly over-predicts the number of lymphoid cancer deaths actually observed in the NIOSH study at the 5% significance level for the cohort overall and for every exposure quintile. In contrast, the predictions of the number of lymphoid cancer deaths by the TCEQ standard Cox proportional hazards model are reasonably accurate and consistent with the number of lymphoid cancer deaths observed in the NIOSH study for the overall cohort and for every exposure quintile.

In addition to this and a demonstration of how USEPA incorrectly calculated model fit criteria in a manner inconsistent with SAB comments, the DSD also presents other new data and evidence not considered by the SAB: important new data on the air concentrations corresponding to endogenous levels (Kirman and Hays 2017), new analyses demonstrating the over-prediction of background lymphoid cancers in the general US population, and robust data indicating the lack of a healthy worker effect for lymphoid cancer and breast cancer incidence (discussed in other comments below). Recent meta-analyses of the association between EtO and human cancer have also been added (Marsh et al. 2019, Vincent et al. 2019). The DSD quite plainly states that as USEPA acknowledges, MOA data do not support their model. Moreover, all other considerations discussed in the DSD also support TCEQ's approach over USEPA's approach (e.g., data on endogenous, reality checks on the key underlying data and background rates in the general population, etc.). Thus, as documented in the DSD, numerous scientific considerations constitute the scientific rationale for departure from USEPA's methodology.

Comment 3:

The DSD incorrectly interprets EPA's statements regarding the plausible sublinearity of dose-response relationships for endogenous doses of EtO as also applying to exogenous exposures.

The DSD states EPA determined "that the low-dose region of the EtO dose-response curve is highly plausibly sublinear…" but this interpretation of the EPA assessment is incorrect. EPA made no such determination about low-dose exogenous exposures.

Rather, EPA made general statements in the context of conceptual models presented by Starr and Swenberg and Crump et al. In this context, EPA was referring to a range of hypothetical endogenous doses from no (zero) endogenous exposure to the point of no (zero) exogenous exposure. The rationale for postulating that the dose-response relationships for relevant cancers across that hypothetical range of doses are likely to be sublinear is based on the conceptual model presented in detail by Crump et al. (2014). In brief, the reasoning is that the body has defense mechanisms (e.g., DNA repair mechanisms) to deal with endogenous exposures. However, these defenses are imperfect and limited, which may account for some level of background cancer risk even without exogenous exposures; and as endogenous doses increase across this hypothetical range, the body's available defenses get diminished, such that the slope of the dose-response curve may be essentially linear at the point of zero exogenous exposure (see Figure 1 in Crump et al. (2014)). The postulated sublinearity is not meant to apply to the range of exogenous exposures.

EPA's unit risk estimates are explicitly for extra risk above background, i.e., above the risk from endogenous doses (unit risk estimates are derived from exposure-response modeling of exogenous exposures; endogenous doses are common to both exposed and unexposed subjects, independent of exogenous exposure, and thus are part of background risk). Variability in levels of background doses of endogenous EtO are accounted for in the modeling of the exogenous exposures, along with other sources of variability. While sublinearity across endogenous doses is plausible, one cannot infer anything from that about the exposure-response relationship at low exogenous exposures. Thus, the DSD's application of the hypothetical sublinear dose-response relationship for endogenous exposures to exogenous exposures, especially in light of background cancer rates (see Comment #4), is not scientifically supported.

Response:

The TCEQ's understanding is that the commenters are indicating that USEPA's unit risk estimates are derived from exposure-response modeling of exogenous occupational exposures and explicitly for extra risk above background, and that endogenous doses were common to both exposed and unexposed workers and thus were part of background risk, independent of the exogenous occupational exposure for which the URF had been derived. The comment further indicates that variability in levels of background doses of endogenous EtO are accounted for in the modeling of the exogenous occupational exposures, along with other sources of variability. However, what the comment fails to consider is that the background for both those exposed and unexposed to EtO at work includes not only endogenous exposure, but

also background exposure to ambient environmental levels outside the workplace in the areas where the workers live (likely reasonably proximate to the facility). Therefore, since endogenous + ambient environmental exposure actually represents background for both those exposed and unexposed at work, *this comment suggests that the URF does not apply to typical environmental levels but only to additional occupational exposure*.

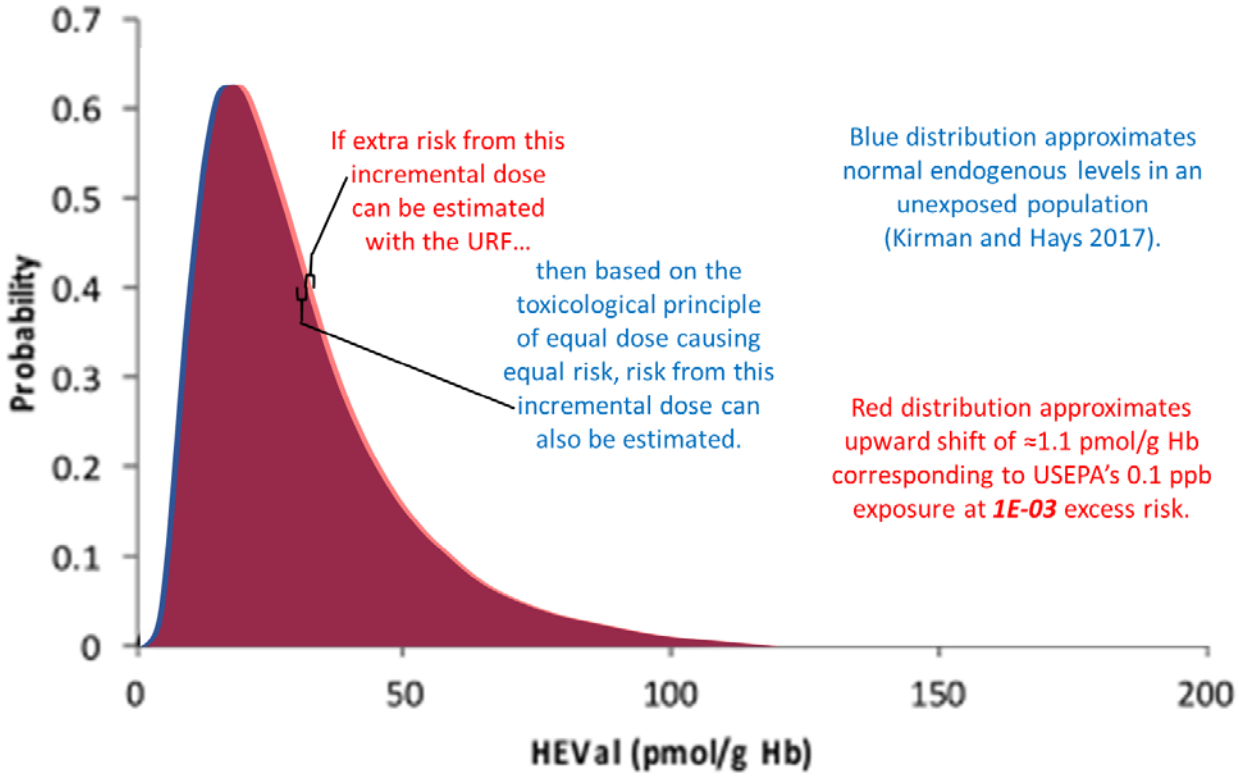


Figure 1. Probability distribution demonstrating the normal endogenous levels of the EtO-hemoglobin adduct in an unexposed population (blue) and in a population with an addition exposure to EtO at 0.1 ppb, associated with 1E-03 excess cancer risk as per USEPA's URF (red)

The basis for the TCEQ's application of the USEPA's URF to understanding risk from endogenous EtO is the toxicological principle that equal internal doses give rise to equal risk. That is, in dose-response/risk assessment it is standard practice to consider equal internal doses as equipotent in producing both noncarcinogenic and carcinogenic effects, absent a scientifically robust demonstration that this standard assumption should not apply (e.g., based on mechanistic or toxicokinetic information). Thus, the standard risk assessment practice of considering equal internal doses as equipotent in producing carcinogenic effects underlies TCEQ's application of the USEPA and TCEQ URFs to internal endogenous doses of the same mutagenic chemical. Internal doses that result from exogenous exposures that fall well within or below endogenous background are expected to be subject to the same defense mechanisms (e.g., DNA repair mechanisms) that deal with endogenous exposures. Thus, the postulated sublinearity still applies to exogenous exposures that result in internal doses within the range of endogenous background, particularly at doses orders of magnitude lower than the 1st percentile of

endogenous as with USEPA's assessment. Based on these toxicological principles, the statement made by the commenter that "the low-dose region of the EtO dose-response curve is highly plausibly sublinear" is a reasonable expectation for the area of the dose-response where total internal exposures (endogenous plus a relatively minor contribution from exogenous) fall within or extremely close to the range of normal endogenous background. This concept has been clarified in the revised DSD. Regardless, it is not deterministic in TCEQ's selection of model. TCEQ used linear low-dose extrapolation from a POD derived by the Cox proportional hazards model, which is indistinguishable from linear over the doses of interest as shown in various figures throughout the DSD (e.g., Figure 22).

Comment 4:

The DSD ignores background rates of cancer and incorrectly suggests that given endogenous EtO production, low exogenous exposures would not produce biologically meaningful internal doses.

The DSD states that "ambient EtO concentrations significantly less than 1 ppb…would not be expected to produce biologically meaningful internal doses considering the range of normal endogenously-produced background EtO levels." However, normal endogenous EtO exposures may contribute to background cancer risks for lymphoid cancers and for breast cancers in females, as these are relatively common cancer types in the general population. As cited on p. 4-95 of EPA's assessment, lymphoid cancers have a background lifetime incidence risk on the order of 3%, while the background lifetime incidence risk for breast cancer in females is on the order of 15%.

Low exogenous EtO exposures would be additive to the endogenous exposure and to background cancer processes, consistent with general principles of quantitative risk assessment. As to the variability in background doses of endogenous EtO, this is accounted for in the modeling of the exogenous exposures, as discussed in Comment #3 above. Thus, DSD ignoring low levels of exogenous exposure claiming they are not biologically meaningful is not scientifically justified and results in an underestimation of risk. For example, the DSD ignores levels of exogenous exposure that EPA determined to be associated with upper bound extra risks of 10^{-4} (0.01%).

Response:

The TEQ stands by the statement that "ambient EtO concentrations significantly less than 1 ppb…would not be expected to produce biologically meaningful internal doses considering the range of normal endogenously-produced background EtO levels." An internal EtO dose that is almost 40 times lower than even the 1st percentile of normal endogenous EtO levels (using USEPA's 0.01 ppb at 1E-04 risk*)* is very unlikely to be biologically meaningful as it is entirely within/below normal biological variation, indistinguishable from background endogenous exposure and risk. Further discussion of the TCEQ's evaluation of endogenous EtO levels is found in response to Comment #3.

The DSD directly addresses background rates of cancer by performing a reality check of USEPA's URF based on them. The TCEQ tested the biological plausibility of USEPA's model and URF and found it both: (1) over-predicted lymphoid cancers in the key NIOSH cohort (based only on exogenous EtO exposures); and (2) over-predicted U.S. population background lymphoid cancer rates based on population-weighted background EtO levels in nonsmokers and smokers.

Comment 5:

The DSD incorrectly attempts to estimate the cancer risks of endogenous EtO levels using EPA's unit risk estimate which is applicable to exogenous exposures only.

The DSD applies EPA's unit risk estimate to endogenous ethylene oxide exposures, but as noted above (Comment #3, 4), EPA's unit risk estimates are for exogenous exposures only (extra risk above background[21]) and cannot be used to infer anything about risks from endogenous exposure. The extent of cancer risks from endogenous levels of EtO is not something that can be estimated from current knowledge.

Response:

This comment is addressed in response to Comment #3. To summarize our response, the background EtO exposure in the workers included both endogenous production as well as ambient EtO exposure. Therefore, it seems that the commenter is suggesting that the URF does not apply to typical environmental levels but only to additional occupational exposure. However, based on the toxicological principle that equal internal doses produce equal risk, the TCEQ considers it reasonable to apply the risks estimated based on occupational exposure to equivalent doses from endogenous or ambient environmental exposure to EtO (i.e., equal internal doses, whether derived from endogenous production or exogenous exposure, should be assumed to be equipotent toxicologically).

Comment 6:

The DSD makes flawed claims about EPA's use of a two-piece spline model for lymphoid cancer and misstates the exposure range over which the model is applied for derivation of the unit risk estimate.

The DSD claims that the EPA model over-predicts the NIOSH cohort results. However, TCEQ's approach to predicting cases is flawed (see Comment #13 below for a discussion of problems in the TCEQ's approach). In fact, EPA's model provides a reasonably good representation of the NIOSH data, as demonstrated by the statistical and visual fits. As seen in Figure 4-3 of EPA's assessment, the model actually underestimates the categorical relative risks (RRs) determined nonparametrically (i.e., without any assumptions about the exposure-response relationship across the exposure categories) for the exposure quartiles.

The DSD claims that EPA was wrong to use a supralinear model. However, the underlying data exhibit a supralinear exposure-response relationship. This is demonstrated by the shape of the nonparametric categorical results, as well as by the fact that the best-fitting models are

supralinear (e.g., the Cox regression model with log cumulative exposure; see Table 4-6 of EPA assessment).

Furthermore, EPA's independent SAB endorsed the use of two-piece spline models for such data, recognizing the utility of such models for reflecting local behavior in the data more readily than the single-parameter models. In fact, EPA used a two-piece linear spline model to account for high-exposure plateauing while specifically avoiding the excessive supralinear curvature in the lower-exposure range objected to by TCEQ and sources it cites regarding supralinear models.

A mechanistic explanation for overall supralinear exposure-response relationships in the observable range of the EtO epidemiological data may not be known; however, such relationships are not uncommon with epidemiological data and there are other possible explanations. Moreover, after modeling all of the data using the two-piece spline model, EPA estimated a point of departure (POD) at the low end of the observable range and used linear low-exposure extrapolation from the POD to derive the unit risk estimate, consistent with EPA's guidelines. (See Comment #11e below for more discussion of EPA's approach to deriving unit risk estimates.) The conclusion of a mutagenic MOA, which was a finding of both EPA and the TCEQ, provides support for linear low-exposure extrapolation. Contrary to intimations by TCEQ, the mutagenic MOA does not preclude high-exposure plateauing, such as exemplified by tumors in rats exposed to vinyl chloride.

Similarly, the plausibility of sublinearity in the conceptual range of endogenous exposures from internal doses of zero up to the point of zero exogenous exposure does not rule out the models used by EPA for exogenous exposures, i.e., supralinearity in the observable range from higher exposures and linear extrapolation for lower exposures.

Thus, the DSD's rationales for rejecting the model used by EPA are not valid.

Response:

The TCEQ agrees that a mutagenic MOA provides support for linear low-exposure extrapolation. The TCEQ also agrees with the commenter that a mechanistic explanation for an overall supra-linear exposure-response relationship is not known, which is the primary (but not sole) basis for TCEQ rejecting USEPA's model (see the DSD). The scientific burden of proof for choosing a supra-linear dose-response is *robust MOA data*, including data over the endogenous range where both agencies expect sublinearity. This burden of proof was not met, and so the TCEQ adopted a standard default dose-response approach (Cox proportional hazards model). Using a standard dose-response model is further strongly supported by the corrected AIC and p-values, which show that the supra-linear model does not have a better fit than the standard model (see Appendix 4 of the revised DSD). The MOA, model predictions of the underlying worker cancer data (Appendix 2 of the revised DSD), and other considerations (e.g., endogenous data) also strongly support the TCEQ's model selection.

The commenter cites a different chemical, vinyl chloride, as an example of a mutagenic carcinogen with high-exposure plateauing of cancer responses in animal models. For EtO, since lymphoid cancer drove the USEPA carcinogenic assessment, perhaps the most relevant mutagenicity data discussed by USEPA (2016) was that in the bone marrow of mice exposed to EtO by inhalation *in vivo* (Recio et al. 2004), which *USEPA indicates is consistent with a linear dose-response* (see C-17 of USEPA 2016). Even at 200 ppm the dose-response remains linear and has not plateaued, making this effect less likely to be occurring in humans.

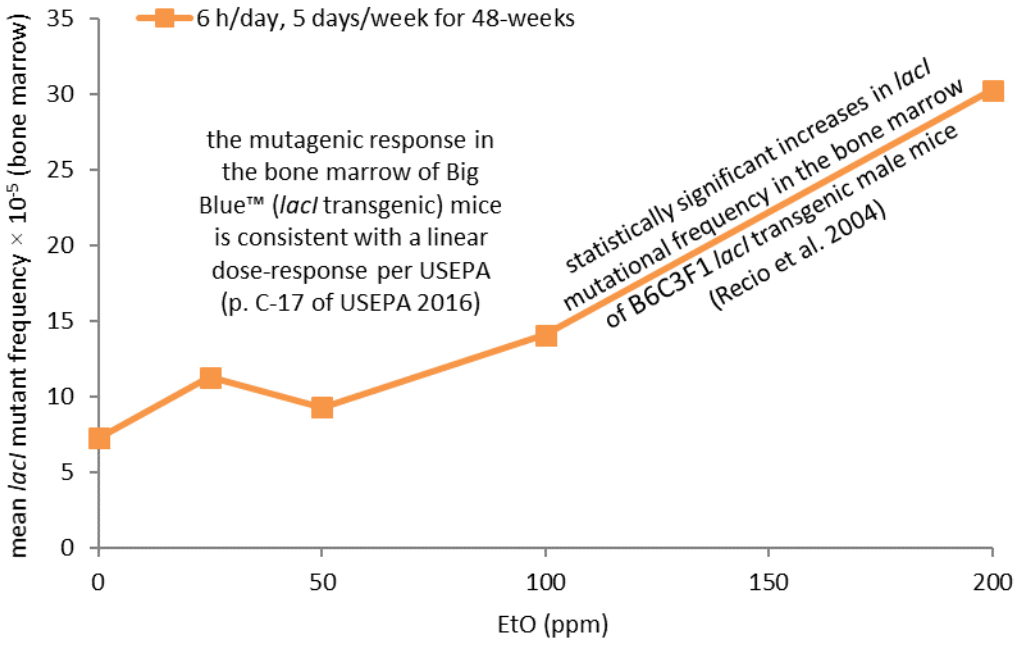


Figure 2. Overall linear dose-response for EtO-induced mutations in the bone marrow of Big Blue™ mice (Recio et al. 2004)

In reference to Figure 4-3 of USEPA's risk assessment, there is an important note at the end of the Figure that addresses model fit: "(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.)" The misinterpretation of model fitting versus non-parametric rate ratios was discussed at length by Valdez-Flores et al. (2013), and USEPA (2016) responded by adding the note to the figures showing rate ratios. There are several reasons why this USEPA Figure 4-3 and other similar figures could be misleading to casual readers. As the models were fit to the individual data and not categorical data, a comparison of models to the categorical data do not actually show model fit at all since those are not the data to which the models were fit. A fair comparison to the individual RRs is shown in Figure 22 of the DSD, which TCEQ therein shows are consistent with the categorical RRs (see Appendix 5 of the revised DSD) but reveals no readily apparent superior fit by either model. This finding using a visual fit method is consistent

with the correctly calculated p-values and AIC values not demonstrating a clearly superior fit by either model. The individual RRs also do not show a high-exposure plateauing effect of EtO. Importantly, TCEQ's reality check on the models predicting the underlying lymphoid cancers for the cohort as a whole and the individual quintiles does show that TCEQ's model, the standard Cox proportional hazards model, more accurately predicts the actual number of lymphoid cancers in the cohort as a whole and in the exposure quintiles. The responses below provide more detailed information on related points.

Non-Parametric Rate Ratios are not Observed Data

Figure 3 reproduces Figure 4-3 in USEPA's 2016 risk assessment. This figure shows the rate ratios of twelve models. Eleven of those models have a parametric functional form and one model (labeled here "EPA's 5 RRs") estimates non-parametric rate ratios of lymphoid mortality grouped by quintiles. Each quintile summarizes information for 11 lymphoid deaths (9 in the non-exposed quintile). *The "EPA's 5 RRs" are not the data* – they are estimates of the rate ratios. Rate ratios are not observed, they are estimated. Furthermore, the non-parametric rate ratios derived by USEPA and shown in Figure 3 do not show the full range of possible rate ratios and cumulative exposures. Figure 4 is identical to Figure 3 with the full range of cumulative exposures of lymphoid decedents and the full range of rate ratios estimated for the individual decedents. Table D-28 of USEPA (2016) includes the uncertainty (i.e., 95% CIs) around USEPA's "categorical" RRs and is reproduced here as Table 2 for lymphoid cancer (males and females combined).

In addition to the categorical RRs not being the actual data or even the data that models were fit to, it must be noted that the apparent supra-linearity of USEPA's 5 RRs are an artifact of the grouping of individual decedents. The functional form implied by non-parametric estimates of the RRs changes depending on the number of lymphoid cancer decedents summarized in each RR. The individual RRs in the figures to follow show the whole spectrum of the RRs and do not suggest a specific shape of the relationship between RRs and cumulative exposures.

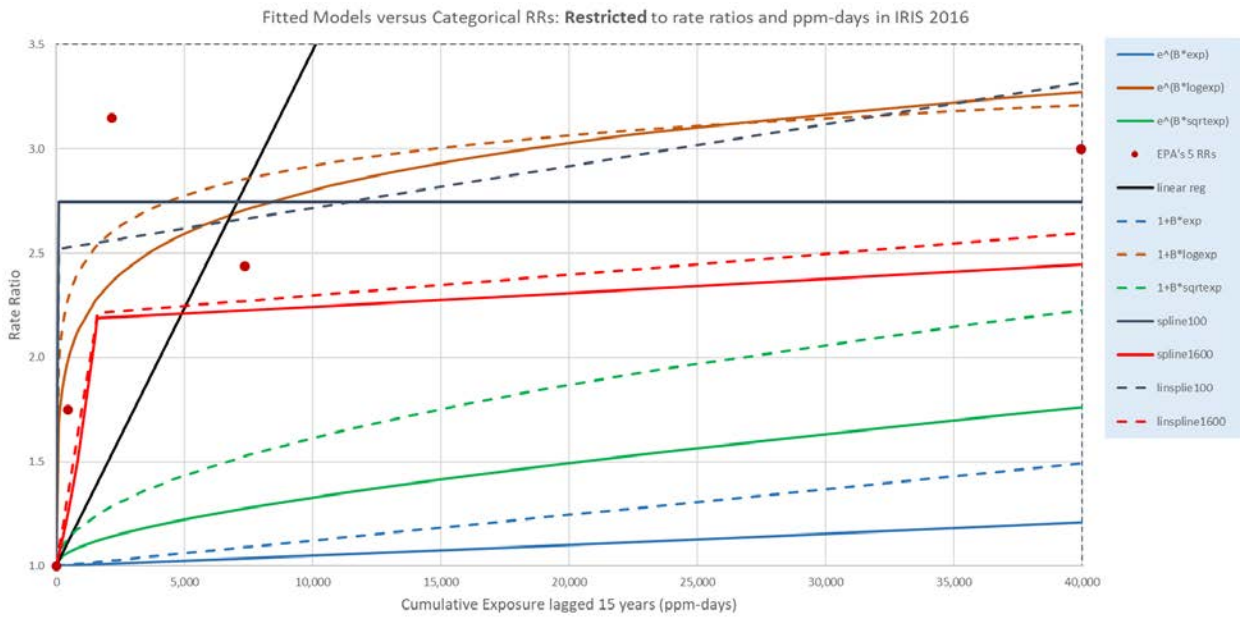


Figure 3. Reproduction of USEPA's 2016 Figure 4-3. Exposure-response models for lymphoid cancer mortality vs. occupational cumulative exposure (with 15-year lag). Spline1600 (solid red line) is the EPA's chosen model, and e^(β*exp) (solid light blue line) is the TCEQ's chosen model.) Spline1600 (solid red line) is the EPA's chosen model, and e^(β*exp) (solid light blue line) is the TCEQ's chosen model.

e^(β*exp): RR = e(β × exposure); e^(β*logexp): RR = e(β × ln(exposure)); e^(β*sqrtexp): RR = e(β × sqrt(exposure)); categorical: RR = e(β × 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.)** [Original footnote with emphasis added]

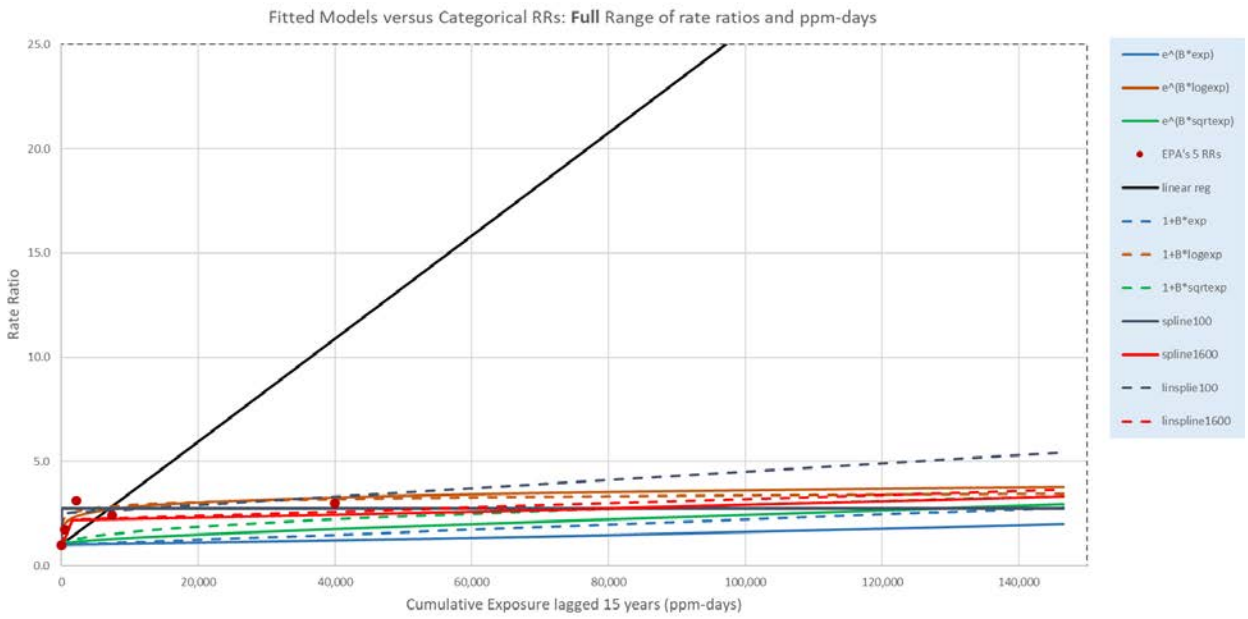


Figure 4. Same as Figure 3 plotted using the full range of cumulative exposures experienced by lymphoid decedents (horizontal axis) and the full range of possible rate ratios estimated for all lymphoid decedents [Note, the choice of y-axis values is based on comparison to Figure 5, discussed below].

Table 2. USEPA's 5 RRs and the 95% confidence interval for each of the quintiles

Cumulative exposure range, 15-year lag (ppm-days)	Mean* Cumulative Exposure (ppm-days)	Rate Ratio	Lower Confidence Limit on the Rate Ratio	Upper Confidence Limit on the Rate Ratio
0 (lagged out)	0	1.00	--	--
>0 – 1,200	446	1.75	0.59	5.25
1,201 – 3,680	2,143	3.15	1.04	9.49
3,681 – 13,500	7,335	2.44	0.80	7.50
>13,500	39,927	3.00	1.02	8.45

* Mean cumulative exposures as reported by USEPA in footnote 21 on page 4-14 of the 2016 assessment.

Estimated Non-Parametric Rate Ratios for Each Lymphoid Decedents

Judging a model fit by comparing to five summary RRs is tantamount to comparing a linear regression fit to 100 different points by comparing it to five averages of groups of 20 points each. Variability of data and goodness of fit of a model cannot be judged by comparing model predictions to five summary points. Similarly, model fit to individual data with 53 decedents cannot be judged by comparing it to five summary RRs. Just as USEPA estimated the non-

parametric RRs by grouping 11 decedents in each exposure quintile, the non-parametric RRs can be estimated for each individual decedent with non-zero exposure (i.e., by fitting a non-parametric hazards model using the Cox proportional hazards model with the RR for each individual adjusted for the covariates sex, race, etc.). The NIOSH study has 44 lymphoid decedents with exposures greater than zero (9 lymphoid decedents had zero lag-15 cumulative exposures). Figure 5 is similar to Figure 4 with the addition of the non-parametric RRs for the 44 lymphoid decedents with non-zero cumulative exposure in the NIOSH study.

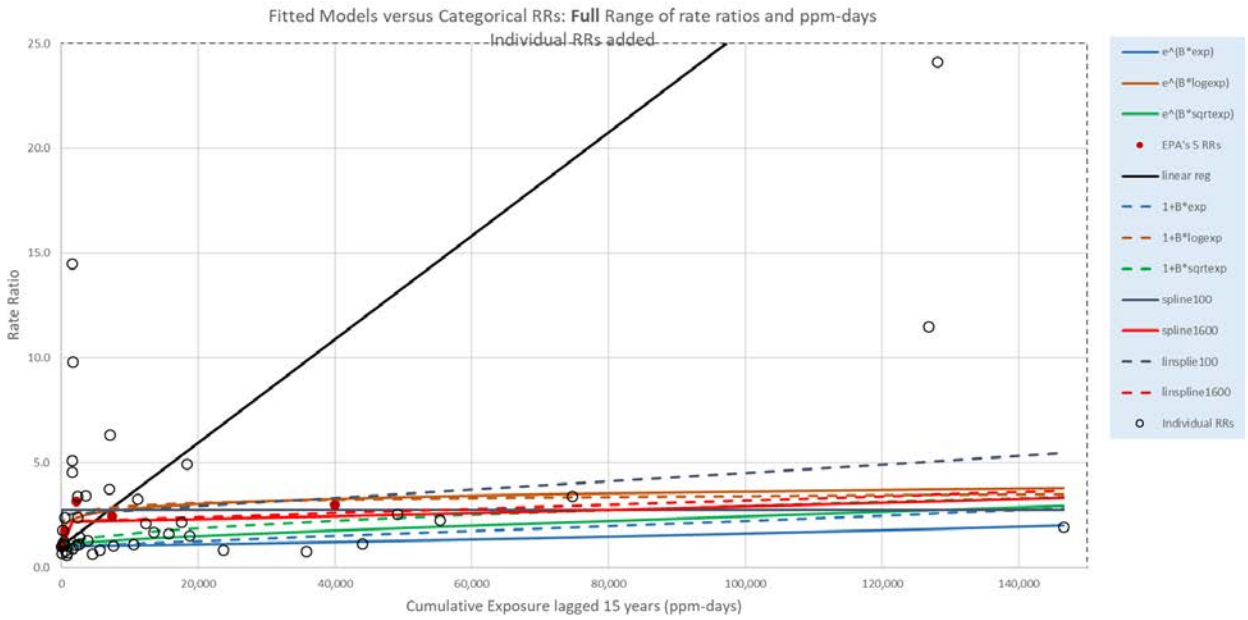


Figure 5. Same as Figure 4 after superimposing the rate ratios estimated for each one of the 44 decedents with non-zero cumulative exposures lagged 15 years.

Figure 5 shows that the non-parametric estimates of the RRs based on the individual decedents can be less than, equal to, or greater than one. This is in contrast to the summary RRs based on quintiles (red dots in Figures 3 to 5) that give the impression that the RRs are greater than one for any non-zero cumulative exposure. Figure 5 also shows that once all the individual variability is included, all models (excluding the "linear reg" model – black solid line) seem to fit the non-parametric estimates similarly. This is supported by correctly calculated p-values and AIC values (Appendix 4 of the revised DSD), although model performance in predicting the actual number of lymphoid cancers in the cohort as a whole and in each quintile demonstrates the superior fit of the Cox proportional hazards model (Appendix 2 of the revised DSD).

Model-Specific Implicitly Estimated Baseline Risks

USEPA's footnote to several figures (including Figure 4-3) indicates that the different models and the non-parametric RRs cannot be compared along the y-axis because "the different models have different implicitly estimated baseline risks." USEPA is correct. All models in Figure

4-3 of the USEPA 2016 EtO risk assessment (Figure 3 herein), with the exception of the "linear reg" model, are fit to hazard rates (not fit to rate ratios). The functional form of all models is

$$HR_i(d) = HR_i(0) \times f_i(d)$$

where $HR_i(d)$ is the hazard rate of model i at cumulative exposure *d,* $HR_i(0)$ is the "estimated baseline risk" for model i, and $f_i(d)$ is the function of the rate ratio at cumulative exposure *d* for model i.

Note that by dividing $HR_i(d)$ by the "estimated baseline risk" $HR_i(0)$, the function $f_i(d)$ is the rate ratio at cumulative exposure *d* for model i. Note also, that each model i could result in different estimates of the baseline risk, $HR_i(0)$. That means, all models would have rate ratio ($f_i(0)$) equal to 1 at cumulative exposure equal to 0. However, the "estimated baseline risk" $HR_i(0)$, could be very different for each model. The model for USEPA's 5 categorical RRs, USEPA's two-piece linear spline model, and TCEQ's standard Cox proportional hazards model have the following functional forms:

Model 1 ("EPA's 5 RRs" and "Individual RRs" in the figures): The non-parametric model fit to the data is given by the expression

$$HR_{NP,k}(d) = HR_{NP}(0) \times RR_{NP,k}(d)$$

where $HR_{NP,k}(d)$ is the hazard rate for the *k*-th group at mean cumulative exposure *d*, $HR_{NP}(0)$ is the "estimated baseline risk" for the nonparametric model, and $RR_{NP,k}(d)$ is the rate ratio for the *k*-th group. Although the function does not depend on the magnitude of the exposure *d,* the function is written with the *d* for the sake of consistency. (USEPA expresses the function $RR_{NP,k}(d) = e^{\beta_k \times d}$ where "*d*" is a "categorical exposure." Using USEPA's expression guarantees $RR_{NP,k}(d)$ is non-negative when doing a search for the optimal β_k values.)

Model 2 ("linspline1600" in the figures): The functional form of USEPA's selected two-piece linear model (linspline1600) is

$$HR_{spl}(d) = HR_{spl}(0) \times \begin{cases} 1 + \beta_1 \times d & d \leq knot \\ 1 + \beta_1 \times d + \beta_2 \times (d - knot) & d > knot \end{cases}$$

where $HR_{spl}(d)$ is the hazard rate at cumulative exposure *d*, $HR_{spl}(0)$ is the "estimated baseline risk" for the two-piece linear model, $1 + \beta_1 \times d$ is the rate ratio at cumulative exposures *d* below the *knot*, $1 + \beta_1 \times d + \beta_2 \times (d - knot)$ is the rate ratio at cumulative exposures *d* above the *knot*, and *knot* is the cumulative exposure where the slope of the rate ratio changes. USEPA estimated *knot* at 1,600 ppm-days from the data.

Model 3 ("e^(β*exp)" in the figures): The functional form of TCEQ's selected model ($e^{\beta * exp}$) standard Cox proportional hazards model is

$$HR_{Cox}(d) = HR_{Cox}(0) \times e^{\beta \times d}$$

where $HR_{Cox}(d)$ is the hazard rate at cumulative exposure *d*, $HR_{Cox}(0)$ is the "estimated baseline risk" for the standard Cox proportional hazards model at zero exposure, and $e^{\beta \times d}$ is the rate ratio at cumulative exposure *d*.

The rate ratios from each of the models described above are, by definition, equal to one at zero cumulative exposures. However, as indicated by USEPA's 2016 assessment and shown above for Models 1, 2, and 3, the "implicitly estimated baseline risks" ($HR_{NP}(0)$, $HR_{spl}(0)$, and $HR_{Cox}(0)$, for Models 1, 2, and 3, respectively) are different for different models. That is, the RRs for the models cannot be compared for non-zero cumulative exposures without accounting for the differences in the "implicitly estimated baseline risks" ($HR_{NP}(0)$, $HR_{spl}(0)$, and $HR_{Cox}(0)$). The partial likelihood methodology for the proportional hazards models described above do not explicitly estimate the baseline risks ($HR_{NP}(0)$, $HR_{spl}(0)$, and $HR_{Cox}(0)$) and they are unknown. However, an approximation of the ratio of the "implicitly estimated baseline risks" for Models 2 and 3 to the "implicitly estimated baseline risks" for Model 1 ($HR_{spl}(0)/HR_{NP}(0)$ and $HR_{Cox}(0)/HR_{NP}(0)$, respectively) can be estimated from the non-parametric RR's based on the individual lymphoid decedents (open circles in Figure 5).

<u>Adjusting Models for Differences in Implicitly Estimated Baseline Risks for More Appropriate Visual Comparison</u>

The ratio $HR_{spl}(0)/\ HR_{NP}(0)$ for Model 2 and $HR_{Cox}(0)/\ HR_{NP}(0)$ for Model 3 were calculated using weighted least squares and the corresponding RR functions for models 2 and 3, respectively. The best intercepts (ratios of baseline risk for each of the models to the baseline risk implied by the non-parametric RR estimates) multiply the rate ratio functions for Models 2 and 3. These adjusted Models 2 and 3 account for the differences in the baseline risks implied by the models and the implicitly estimated non-parametric baseline risks.

Figure 6 adjusts the two-piece linear spline model (linspline1600) and the standard Cox model (e^(β*exp)) by the estimated ratios $HR_{spl}(0)/\ HR_{NP}(0)$ and $HR_{Cox}(0)/\ HR_{NP}(0)$, respectively. These adjusted plots are more appropriate to compare.

The y-axis in Figure 6 has been re-labeled to indicate that the models are normalized to the baseline risk implied by the non-parametric model rather than the models' own implied baseline risks. Figure 6 is divided into four regions using different colors. Each color shows the range of individual RRs" and range of cumulative exposures that are summarized in each of "EPA's 5 RRs" based on quintiles.

[That is, the RR for the highest quintile of "EPA's 5 RR" (red dots) is equal to 3 and is placed at a cumulative exposure of 39,927 ppm-days. Table 2 and Figure 6 show that the RR for the fifth quintile summarizes the individual RRs for the 11 lymphoid cancer decedents (open circles) that had cumulative exposures greater than 13,500 ppm-days. Similarly, the RR for the fourth quintile summarizes the 11 individual RRs (open circles) based on lymphoid decedents with cumulative exposure between 3,681 and 13,500 ppm-days. The RR for the third quintile summarizes the 11 individual RRs (open circles) based on lymphoid decedents with cumulative

exposure between 1,201 and 3,680 ppm-days. Finally, the RR for the second quintile summarizes the 11 individual RRs (open circles) based on lymphoid decedents with cumulative exposure greater than zero and less than or equal to 1,200 ppm-days.]

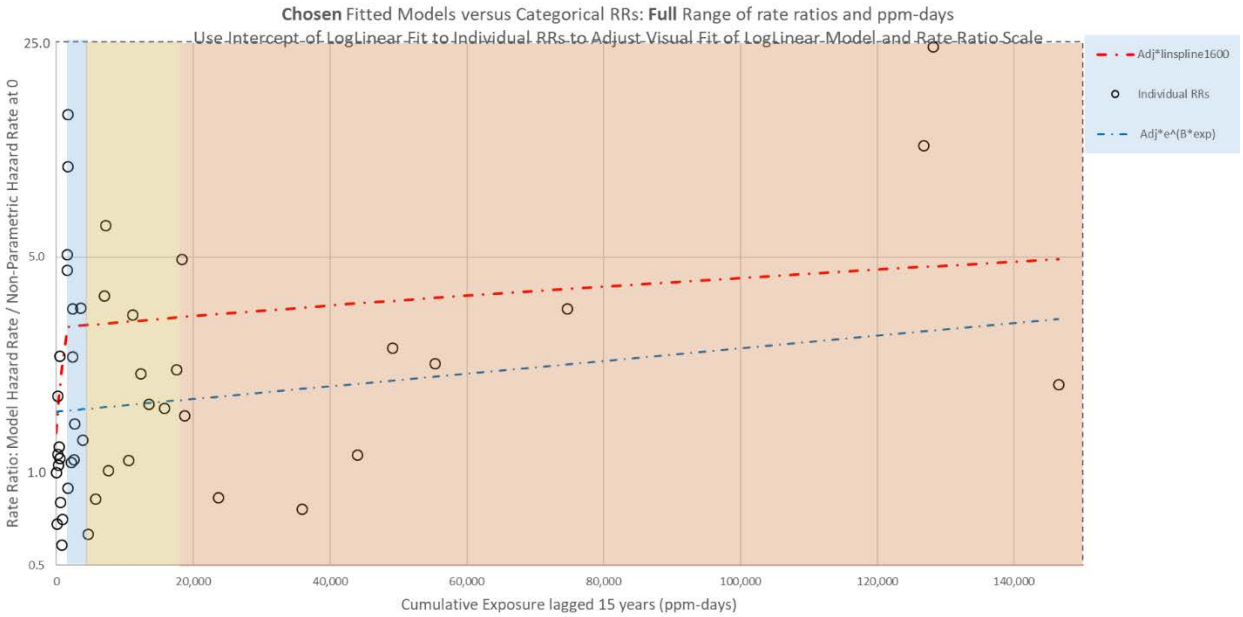


Figure 6. USEPA's and TCEQ's fitted models after adjusting the models for the difference between the implied background hazard rates of the models and the non-parametric background hazard rate compared to the RRs based on the individual decedents.

Figure 6 shows that the model selected by USEPA ("linspline1600") cannot be visually judged to be better than TCEQ's model ("e^(β*exp)").

In summary, although a secondary consideration to statistical analyses, visual comparisons of USEPA and TCEQ selected models fit the individual RRs approximately the same once differences in "baseline risks" of different RR models are reconciled. This conclusion is consistent with the conclusions drawn using correctly calculated standard model fit criteria (Appendix 4 of the revised DSD, also described in response to Comment #7 below). Moreover, the TCEQ's assessment of model performance at predicting the actual number of lymphoid cancers in the cohort as a whole and in each quintile demonstrates that the Cox proportional hazards model is reasonably accurate while USEPA's model is statistically significantly over-predictive (see Appendix 2 of the revised DSD).

Comment 7:

The DSD criticism of how EPA addressed the knot in the two-piece spline models are contrary to SAB recommendations to the EPA.

The DSD objected to the fact that EPA did not include the knot as a parameter in its estimations of the Akaike Information Criterion (AIC). Inclusion of the knot as a parameter would have been one way to do the calculation; however, the SAB supported EPA's approach. Consistent with SAB recommendations, the EPA did not make its model selections based solely on the AICs. As recommended by the SAB, EPA also gave consideration to the ability of models to reflect local behavior, e.g., prioritizing two-piece spline models, and to parsimony. The SAB singles out the knot as a parameter that could be fixed in the interest of parsimony, stating "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…. In the draft assessment, fixing the knot when estimating linear spline model fits from relative risk regressions is one such example." Moreover, the SAB fully understood how EPA determined the knot, having reviewed the Agency's approach as a charge question, finding it "scientifically appropriate and a practical solution that is transparently described." Thus, the DSD objections that the knot was estimated before it was fixed are not persuasive.

Response:

This comment addresses three different points related to the SAB review of EPA's EtO modeling: 1) whether the USEPA transparently described how the knot was set; 2) whether the knot should be considered to be a "fixed" parameter (and therefore not contribute to the number of degrees of freedom in the AIC and p-value calculations), or if the knot should be considered to have been estimated from the data, and therefore should contribute to the number of degrees of freedom; and 3) prioritizing models that reflect local behavior (such as spline models) and model parsimony.

Addressing the first point, the commenter refers to the following from the 2015 SAB document (p. 13), stating that "the SAB fully understood how EPA determined the knot" [*emphasis added*]:

> 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 TCEQ agrees that the method that was used to identify the knot was clear. However, how USEPA determined the knot ("the value at which the likelihood is maximized") is an entirely separate issue from whether USEPA appropriately accounted for the methods used in determining their degrees of freedom.

Relevant to the second point, the SAB stated that "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" (p. 12 of SAB 2015). But as noted above, the knot values were statistically estimated/optimized based on the NIOSH data, so did not conform to the SAB's condition that some model parameters may be fixed in the interest of parsimony if they are not estimated from the data. Therefore, the TCEQ's conclusion that the knot derivation (which was estimated from the data) should be counted towards the degrees of freedom is not contrary to the SAB's recommendations that parameters that are not estimated from the data should be fixed in the interests of parsimony. More detailed discussion is provided below.

Model-fitting exercises were used in setting the knot values. Under the commenter's point of view, multiple parameter estimates from upstream model-fitting exercises could be used to "fix" a downstream model's parameter values (at a midpoint or mean of the estimates, for example), yet none of them would count as fitted parameters. Estimating parameters from the data, and then using the parameters as if they were not estimated is not statistically appropriate. Tables in the USEPA (2016) risk assessment and its appendices clearly indicate that the knot of the two-piece spline models were determined so that the likelihood was maximized. For example, on page 4-23 USEPA states, "For this assessment, the knot was generally selected by evaluating different knots in increments of 100 ppm × days over some range of cumulative exposures starting at 0 and then *choosing the one that resulted in the best (largest) model likelihood*. The model likelihood did not change much across the different trial knots for any of the data sets, but it did change slightly, *and the largest calculated likelihood was used as the basis for knot selection*." [*emphasis added*]. However, USEPA uses p-values and AIC values that do not account for the fact that the knot is being estimated. This is evident in Table D-30 of USEPA's appendices, which lists the p-value of 0.0482 and an AIC of 461.847. These numbers (with some rounding) are listed in USEPA's Table 4-6. What must be appreciated is that the number of degrees of freedom (DF in Table D-30, shown in Figure 7 below) is equal to 2 (one for LIN_0 and one for LIN_1). The table comes from an SAS run output that correctly accounts for the estimated parameters in the model being fit. However, SAS does not account for the fact that the output of Table D-30 is selected because it has the largest likelihood (smallest deviance) with covariates amongst all models with a knot. The number of DF reported in the SAS output (2), therefore, does not account for the fact that that the likelihood corresponds to the model whose knot maximizes the likelihood. To account for the fact that the model was selected so that the likelihood was maximized, USEPA would have to calculate the p-value and the AIC value externally. Instead USEPA quoted the p-value and AIC value as reported by the SAS output, which does not account for the fact that that the knot was estimated from the data to maximize the likelihood (inconsistent with the SAB's recommendation of a fixed parameter not estimated from the data). This error in the calculation of p-values and AIC values occurred for all two-piece models reported in the USEPA (2016) risk assessment.

The TCEQ DSD corrected these calculations in the USEPA (2016) risk assessment and reported the correct calculations in Appendix 4 of the revised DSD.

Table D-30. Results of two-piece log-linear spline model for lymphoid cancer mortality, men and women combined, knot at 100 ppm-days

Model fit statistics					
Criterion	Without covariates	With covariates			
-2 LOG L	463.912	457.847			
AIC	463.912	461.847			
SBC	463.912	465.787			
Testing global null hypothesis: BETA = 0					
Test	χ^2	DF	Pr > ChiSq		
Likelihood ratio	6.0658	2	0.0482		
Score	5.9648	2	0.0507		
Wald	5.8246	2	0.0544		
Analysis of maximum likelihood estimates					
Parameter	Parameter estimates	Standard error	χ^2	Pr > ChiSq	Hazard ratio
LIN_0	0.01010	0.00493	4.1997	0.0404	1.010
LIN_1	-0.01010	0.00493	4.1959	0.0405	0.990

Figure 7. Results of two-piece log-linear spline model for lymphoid cancer mortality, men and women combined, know at 100 ppm-days (USEPA 2016)

The third point in the comments indicate, "As recommended by the SAB, EPA also gave consideration to the ability of models to reflect local behavior, e.g., prioritizing two-piece spline models, and to parsimony." Although those words were taken from the SAB review, there is greater clarity in the full quote from the SAB:

> "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."

The first SAB recommendation was taken into account by the USEPA (2016) risk assessment by using the individual data. USEPA also considered the second SAB recommendation. However, the SAB's second recommendation refers to prioritizing spline models over models with transformation of exposures (e.g., the model with log-transformed cumulative exposures). *The second recommendation is **not** prioritizing spline models over all other models*, only over models with transformation of exposure (e.g., square root of cumulative exposure, log of cumulative exposure, etc.). If the second recommendation were about prioritizing the spline model over all other models, then the SAB would be contradicting themselves in the third recommendation that appeals to prioritizing parsimony (i.e., preferring simpler models).

Comment 8:

The DSD inappropriately uses a Cox proportional hazards (PH) model for the NIOSH cohort, despite its lack of fit to the data.

a. As a central part of its analysis to calculate the cancer unit risk estimate, the DSD uses a model for the NIOSH cohort that they note does not provide a statistically significant fit to the data (though the DSD does not present a p-value). In addition, the approach that they used to demonstrate that their model provides good "predictions" of the number of cases in the NIOSH cohort is flawed (see Comment #13 below).

b. Furthermore, the model used by TCEQ is inherently sublinear and cannot reflect the overall supralinear shape of the exposure-response relationship (See model "e^(β*exp)" in Fig 4-3 of EPA's assessment and the p-values in Table 4-6). The Cox PH model for lymphoid cancers in males and females in the NIOSH cohort has a p-value 0.22, while the best-fitting supralinear model has a p-value of 0.02—a much lower and statistically significant value, indicating the supralinear model provides a better fit to the data. The Cox PH model was presented in EPA's assessment for comparison with other models, therefore the SAB was able to consider it as an option, and yet, the SAB did not promote it but instead endorsed two-piece spline models.

   EPA and the SAB recognized the importance of local fit to the data, as well as overall fit. The two-piece spline model used by EPA, and endorsed by the SAB, can represent the increasing response at lower exposures (without excessive curvature at the lowest exposures) and the relative plateauing at higher exposures, as discussed above (Comment #6). To estimate the risks of environmental exposure levels from higher exposure data, such as occupational data, capturing this local behavior at the lower exposure range of the data is especially important because it reflects the data range most relevant to the even lower exposures of interest.

c. In contrast, the Cox PH model used by TCEQ cannot accommodate supralinear exposure-response data and, in particular, cannot reflect the exposure-response relationship in the lower exposure range of the data. Instead, in order to attempt to fit the high-exposure plateauing, such a model must inflate the internal baseline hazard rate and depress the low-exposure slope. This is illustrated in Figure 21 of the DSD, where the dotted blue line depicts the model used by TCEQ with an approximated baseline hazard rate shown relative to the nonparametric baseline hazard rate. It is apparent from this depiction that the baseline rate in the Cox PH model has been markedly overestimated relative to the nonparametric (categorical) baseline (RR = 1). The nonparametric baseline, however, is the best available estimate of the baseline hazard rate in the cohort because it is based on the 0 (lagged) cumulative exposure group without any assumptions about the shape of the exposure-response model for the exposed workers and, thus, without any influence of the higher-exposure data on the model fit to the lower-exposure data.

d. Despite these clear problems, TCEQ goes on to calculate the point of departure (POD) using the sublinear Cox PH model, and Table 30 on p. 93 of the DSD presents a confidence score of "high" for the POD. This score is totally unwarranted because the Cox PH model does not provide a statistically significant fit to the data and is inconsistent with the overall shape of the exposure-response data. Moreover, the "predictive" value of the model is based on a flawed approach, as discussed in Comment #13 below. Finally, even if the model and predictions were valid, there are insufficient data in the range of the POD, which was calculated at a risk level of 1 in 100,000, to conclude that the model yields reliable estimates in that range, as discussed in Comment #11.

For all of these reasons, the DSD's model selection and POD derivation, and the subsequent cancer unit risk estimates based on them, are not scientifically supported.

Response:

a. Based on correctly calculated p-values and AIC values (discussed in response to Comment #7), the Cox proportional hazards model fits the data as well as USEPA's overall supra-linear model (see Appendix 4 of the revised DSD for correctly calculated p-values and AIC values). That being said, neither TCEQ's model nor USEPA's selected model (two-piece linear spline model with knot at 1600 ppm-days) is statistically significantly different than the model with zero slope once the correct degrees of freedom are correctly accounted for. However, the model TCEQ used is preferable to USEPA's selected model because it has a better AIC than the model USEPA selected (Table 38 in the proposed TCED DSD). The calculation of a unit risk factor is done for regulatory purposes and not because EtO exposures had shown increasing cancer risk in the NIOSH study. Valdez-Flores et al. (2010) fit the model to 72 combinations of 12 endpoints and 6 sub-cohorts, lagged and unlagged, and found none were statistically significantly related to cumulative exposures of EtO; these results are consistent with the recent meta-analyses by Marsh et al. (2019) and Vincent et al. (2019). Lymphoid cancer in NIOSH male workers had the lowest effective concentration found by Valdez-Flores et al. (2010), which the TCEQ conservatively choose to estimate unit risk factors (even though there is no significant increased cancer mortality with cumulative exposure to EtO for any of the endpoints examined). Also see the response to Comment #13.

b. Regarding the comment on the supra-linear shape of the exposure-response relationship, refer to the response to Comment #6 above as well as the following. While visual fit to the nonparametric RRs was used by USEPA (2016) as a criterion for model selection, no appropriate visual comparison of model fit to the lymphoid cancer mortality data can be made based on USEPA Figure 4-3 (p. 4-21 of USEPA 2016) since the data shown are not the data to which the models were fit. USEPA Figure 4-3 shows models compared to arbitrarily-chosen quintile categories of data used to estimate rate ratios. The actual data underlying the model fits are the individual data, not the less refined categorical data shown in USEPA Figure 4-3. Thus, because the model fits shown in USEPA (2016) Figure 4-3 are fit to the individual data (and not the categorical data depicted), the figure does not actually indicate model fit to the modelled data at all. Additionally, categorical non-parametric rate ratios

(RRs) should not be used for visually comparing models fit to individual data, particularly when appropriate statistical model fit criteria are available. The RRs of parametric models fit to the individual data are defined with respect to an underlying background hazard rate estimated by the model. However, the underlying background hazard rates estimated by the nonparametric RRs and the parametric model are generally different. Moreover, visual interpretation of the consistency of categorical RRs with the shape/slope of a modelled dose-response can change as the number of exposure categories changes. For example, Figures 1-3 of Valdez-Flores and Sielken (2013) demonstrate, among other things, how the dose-response (i.e., dose-RR) slope for breast cancer mortality in the NIOSH cohort appears very steep when compared to only four exposure categories but seems more shallow when additional categories are added. In the present case, the overall dose-response appears ill represented by only a few categorical RRs (see Valdez-Flores and Sielken 2013 and associated supplementary materials). The visual presentation of only a few exposure categories can blind the data user to the variability in the underlying dose-response data, and by corollary, preclude an appropriate visual assessment/comparison of model fit to the actual individual data. See Appendix 5 of the revised DSD for additional information. A better comparison of models fit to the observed data is to use the predictiveness of the model; that is, the capability of the model to estimate the observed number of deaths with a certain degree of confidence (see Appendix 2 of the revised DSD).

Regarding the SAB's opinions on model choice, none of the charge questions asked the SAB to make a recommendation as to which model was best. The SAB did however, include a discussion about using the AIC. They indicated that the AIC could be used to compare models as long as the models were not using a transformation of the cumulative exposures (e.g., log(cumulative exposure) or square root(cumulative exposure)). Thus, following the SAB recommendation and using the corrected AIC and p-values from Table 4-6 (as reported in Table 38 of the proposed TCEQ DSD and discussed in response to Comment #7), the TCEQ log-linear model has an AIC of 464.4 and the USEPA selected two-piece linear spline model has an AIC of 464.5 (lower AIC scores are generally preferable). Furthermore, the TCEQ log-linear model is more parsimonious (has fewer parameters) than USEPA two-piece spline models with a knot at 1600 ppm-days. While the SAB did recommend using a two-piece spline model *over the models that use the logarithm of the cumulative exposure or the square root model*, they did not dismiss the Cox proportional hazards model, which based on new analyses by the TCEQ clearly better predicts the observed number of deaths in the cohort overall and in every cumulative exposure quintile (see Appendix 2 of the revised DSD).

c. The commenters note that "the Cox PH model used by TCEQ cannot accommodate supralinear exposure-response data". It is true that the Cox proportional hazards model is not a supra-linear model; however, the conclusion that the data is supra-linear is based (as noted above) on the categorical quintile analysis that does not represent the underlying data. In fact, the best fitting model to the underlying data is a piecewise model that joins the individual RRs (44 individual RRs and the control group), producing a clearly meaningless relationship:

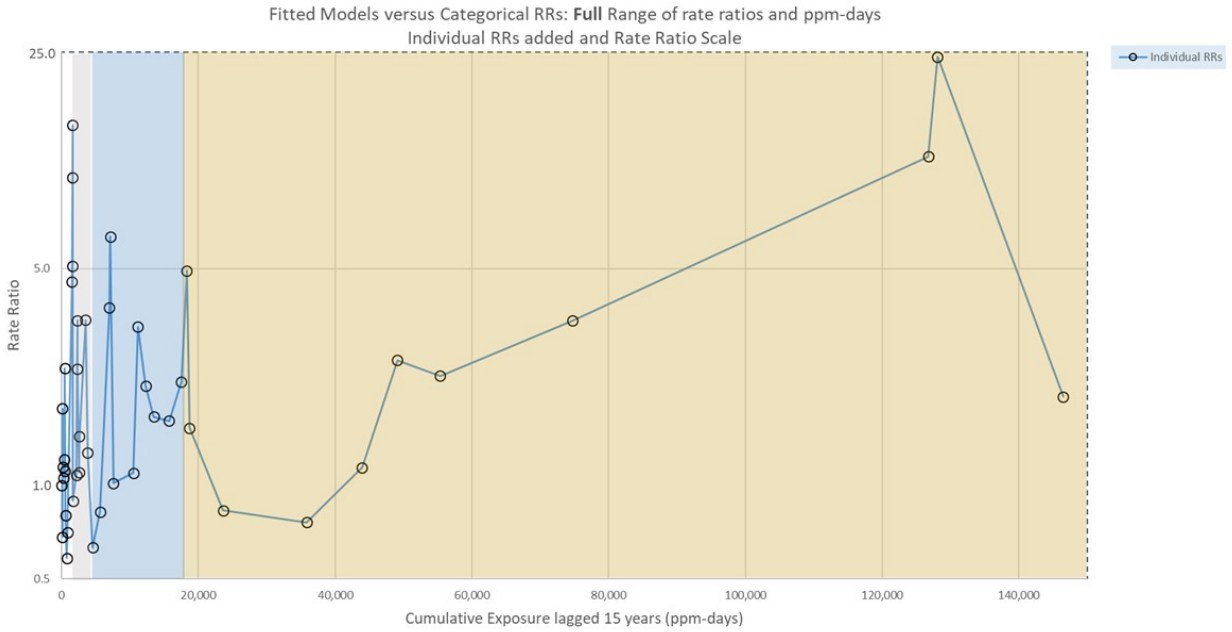


Figure 8. Piecewise model joining the 44 individual RRs and the control group.

The Cox proportional hazards model is a parsimonious model that compromises across the variability present in the data, weighting observations according to the risk set of each individual decedent (53 lymphoid decedents in the NIOSH study). The non-parametric rate ratios are estimated as the ratio of the mean hazard of a group of individuals (the group may include one or more decedent) and the mean hazard of all the individuals in the unexposed group (there are 9 decedents with 0 cumulative exposure when lagged 15-years in the NIOSH study). The mean hazard of the 9 individuals in the unexposed group is not a constant, fixed number. The implicitly estimated mean baseline hazards have variability that is not estimated in the Cox proportional hazards models (the non-parametric rate ratios are one such example). That is, the non-parametric rate ratios do not have the "true" implicitly estimated baseline hazards; they have the implicitly estimated baseline hazards that maximize their likelihood. That is true also for all Cox proportional hazards models (TCEQ and USEPA models) fit to the NIOSH individual data. In summary, an exposure-response model is not expected to go through any specific nonparametric rate ratio, nor is expected to implicitly result in the same baseline risk estimated non-parametrically. TCEQ response to Comment #6 elaborates on this issue and shows figures that may help the reader better understand how Cox proportional hazards models can be visually compared.

d. As indicated above, TCEQ developed a POD despite the fact that the recent EtO literature does not support any increased significant cancer risks in humans (Marsh et al. 2019, Vincent et al. 2019, Valdez-Flores et al. 2010). The calculation of a unit risk factor is done for regulatory purposes and not because EtO exposures had shown increasing cancer risk in the NIOSH study. Furthermore, the standard Cox model estimates the number of observed lymphoid deaths in the NIOSH study (the actual observed data, rate ratios are not observed

data) with 95% confidence. USEPA's selected model is shown by TCEQ to overestimate the observed number of lymphoid cancer deaths at the 5% significance level.

In contrast to the commenter's statement that there are insufficient data in the range of the POD (1 in a 100,000), the TCEQ calculated that 7 of the 27 male lymphoid decedents had cumulative exposures below the POD for 1 in 100,000 extra risk (Table 1). Similarly, more than 30% of the male workers in the NIOSH study were exposed to EtO concentrations below the 1 in 100,000 extra risk. In addition, the URF changes less than 8% if a higher POD (1/1,000 extra risk) is considered. A POD of 1 in 100 (as used by USEPA) results in concentrations that exceed the cumulative exposure for all male lymphoid decedents and the cumulative exposure of more than 99.8% of all male workers in the NIOSH study.

The rest of the comments in this paragraph are addressed in previous or subsequent comment responses.

Comment 9:

The DSD is incorrect in its claim that EPA should have considered environmental exposures to ethylene.

Environmental exposures to ethylene would be part of background risk and would not affect EPA's EtO unit risk estimate, which is for extra risk above background.

Response:

The TCEQ DSD does not state that USEPA (2016) should have considered environmental exposures to ethylene in their dose-response assessment. Rather, the TCEQ is noting that USEPA has not followed the implications of their EtO assessment for ambient ethylene; namely that based on the USEPAs URF for EtO, given the 3% conversion of ethylene to EtO in humans cited by USEPA, there would appear to be a nationwide ethylene carcinogenic risk issue (which the TCEQ does not find is the case).

Comment 10:

The DSD ignores issues with the Swaen et al. (2009) analysis that decreased the ability of that analysis to detect associations for lymphoid cancer.

The DSD cites the Swaen, et al. (2009) study of the Union Carbide Corporation (UCC) as reporting that "no indications were found for excess cancer risks from EtO exposures, including lymphohematopoietic malignancies," however the Swaen analysis has important limitations:

a) The trend analyses were done using the sublinear Cox model, which would be limited in detecting supralinear trends (see Comment #8).

b) The categorical analyses were based on standardized mortality ratios (SMRs), which are notoriously deficient for analyzing occupational epidemiology data because workers

often have background disease mortality rates below those of the general population. This concept is called the "healthy worker effect" (HWE), although it can reflect differences between an occupational cohort and the general population beyond health. In fact, EPA's SAB specifically recommended that epidemiological results based on external standards, e.g., SMRs, be down-weighted, stating "[t]he presence of the healthy worker effect cannot be denied in these occupational data and the use of an external standard for comparison does not avoid healthy worker types of biases."

c) The long follow-up in the UCC cohort, well past the occurrence of non-negligible exposures, was likely observing proportionately more background cases associated with increasing age of the cohort than cases associated with exposures in the distant past. In other words, most of the workers who would die of exposure-related lymphoid cancers would likely have already passed; thus, proportionately more of the new cases picked up in the extended follow-up would be background cases. This excessive follow-up, given the time that had lapsed since non-negligible exposures ceased, would make it more difficult to observe an exposure-related effect. (See also p. A-30 to A-31 of Appendix A of EPA (2016b) for more discussion.)

The DSD's interpretation of the Swaen study does not account for these critical limitations.

Response:

a) Regarding the statement that "The trend analyses were done using the sublinear Cox model, which would be limited in detecting supralinear trends (see Comment #8)", as shown in various figures in the DSD, the Cox model is indistinguishable from linear over the doses of interest for EtO. In addition, the analysis from Swaen et al. (2009) assessed whether the slope of the exposure-response relationship was statistically significantly different from zero; it was not, and does not reflect the shape of the exposure-response relationship.

b) In evaluating internal standard analyses, USEPA's SAB specifically stated, "[t]he presence of the healthy worker effect cannot be denied in these occupational data and the use of an external standard for comparison does not avoid healthy worker types of biases." However, the most pertinent question is whether a healthy worker effect can be assumed for the cancer endpoints being evaluated for EtO. That is, as SAB suggests, is the healthy worker effect for lymphoid cancer and breast cancer incidence in fact undeniable? Though opinions vary about using general population background rates for evaluating cause-specific mortality rates of occupational studies, it is standard practice/methodology in epidemiology studies to use general population background rates to evaluate the well-established and widely-accepted SMRs and SIRs because there is often no scientific evaluation of the magnitude of the healthy worker effect.

   The TCEQ investigated this question and finds that the healthy worker effect (as evidenced by decreased overall mortality, etc.) does not necessarily extend to specific cancers. More specifically, the results of a relatively recent and large study (366,114

workers) conducted specifically to examine the potential for the healthy worker effect in cancer incidence studies (Kirkeleit et al. 2013) found, for example, that while all-cause, ischemic heart disease, and circulatory system disease mortality were statistically significantly decreased in male workers (n=283,002) and female workers (n=83,112) compared to the general population (Table 3 of the study), the SIRs for lymphoid and hematopoietic cancers in male workers and female workers were 0.97 (0.90, 1.03) and 1.09 (0.92, 1.27), respectively, consistent with the lack of a statistical difference (i.e., lack of a healthy worker effect for these cancers). Additionally, the Kirkeleit et al. (2013) study found that breast cancer incidence in over 83,000 female workers was as expected based on the general population (i.e., SIR of 1.02 (0.95, 1.09)). Thus, Kirkeleit et al. (2013) found no healthy worker effect for both lymphoid/hematopoietic and breast cancer incidence, and SMRs for workers in the NIOSH and DOW/UCC are similar to those reported by Kirkeleit et al. (2013) for lymphoid cancer mortality. Footnote * to Table 32 in the TCEQ DSD states "Quintile 1 is the control (unexposed lagged-out) group with 9 lymphoid cancer mortalities observed and 11.5 mortalities predicted by all models with a 95% confidence interval of (6.0 and 25.2), which includes the 9 lymphoid cancer deaths." This confidence interval indicates that the observed 9 lymphoid cancer deaths in the unexposed male and female workers of the NIOSH cohort is consistent with the number of lymphoid cancer deaths in the general U.S. population during the same period of time after accounting for age, sex, and calendar year. Expressed in terms of SMRs, the SMR for lymphoid cancer deaths in the unexposed male and female NIOSH workers is equal to 0.78 (9/11.5) with a 95% confidence interval (CI) equal to (0.36, 1.50). The 95% CI on the SMR for unexposed workers includes the value of one, which indicates that the mortality rate in the unexposed workers in the NIOSH study and the U.S. population mortality rate are not statistically significantly different at the 5% significance level (p-value of 0.29). Similar results are obtained for the male NIOSH workers that drive lymphoid cancer risk. More specifically, the SMR for lymphoid cancer deaths in the unexposed male NIOSH workers is equal to 1.03 (6/5.8) with a 95% CI of (0.38, 2.25). Thus, the lymphoid cancer mortality rate in unexposed male workers in the NIOSH cohort, the gender that drives the URF, is not statistically significantly different than that in the U.S. population (p-value of 0.64).

In summary, these results demonstrate that there is no healthy worker effect for this critical endpoint in this key group (i.e., male workers, who drive lymphoid cancer risk in the cohort) or in males and female workers combined. These results based on the NIOSH cohort are consistent with the findings of Kirkeleit et al. (2013). Thus, the comments do not change TCEQ's overall conclusions on the findings reported by Swaen et al. (2009) using external background hazard rates (which were corroborated by Valdez-Flores et al. (2010) using internal background hazard rates).

c) To address the concerns from sections a, b, and c of these comments, the DSD now refers to section A.2.20 of USEPA (2016), which includes a discussion of Swaen et al. (2009), and language has been added regarding the long follow-up being viewed as a limitation by USEPA, as well as their characterization of the exposure assessment as

relatively crude. The TCEQ notes, however, that the UCC cohort data does not contribute to the final URF (i.e., TCEQ's URF is exclusively based on NIOSH cohort results). In addition, even given the lack of statistically significant association between lymphoid cancer and EtO concentrations in the UCC cohort, the TCEQ conservatively assumes that EtO causes lymphoid cancer in humans.

Comment 11:

The DSD's approach to deriving a quantitative cancer risk estimate for ethylene oxide exposure has a number of scientific problems that lead to underestimating risk.

a) TCEQ's quantitative risk estimates are for lymphoid cancer only and do not include the risks for breast cancer in females (see also Comment #1).

b) For lymphoid cancer, as discussed above (Comment #8), the DSD selected a sublinear Cox PH model that does not fit the data.

c) In addition, the use of a 70-year cut-off in the lifetable analysis is not consistent with a default (average) lifetime of 70 years. EPA also uses a default average lifetime of 70 years but recognizes that 70 years should not be used as a cut-off in lifetable analyses, because in such analyses, actual demographic data about mortality rates at different ages are incorporated rather than using an average default lifetime. Truncating the analysis at 70 years actually corresponds to an average lifetime of less than 70 years because the hypothetical population tracked in the lifetable analysis is allowed to die at younger ages than the would-be average of 70 years but not allowed to live beyond 70 years. In contrast, truncating the lifetable analysis at 85 years corresponds to an average lifetime of about 75 years, which is close to the default average of 70 years.

d) Another difference between the EPA and TCEQ approaches is that the EPA estimates are for cancer incidence, whereas the TCEQ estimates are for mortality. The SAB endorsed EPA's approach for calculating incidence estimates from mortality data.

e) Moreover, in the DSD, the modeling for lymphoid cancer was apparently done all the way down to a risk level of 1 in 100,000 using the (non-fitting) sublinear model. In so doing, the TCEQ over relies on a sublinear model that doesn't describe the overall data well and certainly can't reliably estimate risks at corresponding low levels of exposure where there are few data. In other words, this approach assumes that the sublinear model is valid not only in the observable range of the data, contrary to findings that the underlying exposure-response data are more supralinear in shape, as discussed above (Comment #8), but also in the lower exposure range, where the data are insufficient to estimate risks with any confidence. On p. 5 of the DSD, the TCEQ criticizes the EPA, stating "High-dose carcinogenicity data alone are incapable of informing truly low-dose risk"; however, it is the TCEQ, not the EPA, that models from the high-dose data down to a risk level of 1 in 100,000.

In contrast, EPA's approach does not presume to be able to estimate risks at such low levels. Instead, EPA's Guidelines on Carcinogen Risk Assessment advocate modeling the data and then selecting a POD near the low end of the observable range, i.e., the low end of the range in which increased risks might be reasonably detectable above background variability, and applying an extrapolation method from the POD. In the absence of sufficient evidence that a nonlinear approach is warranted, the default approach is to use linear extrapolation. In the case of EtO, the use of linear extrapolation from the POD is supported by the finding of a mutagenic MOA, in accordance with EPA's guidance. Linear extrapolation was also endorsed by the SAB. Given the background rates of lymphoid cancer, EPA chose a POD of 1% extra risk, or 1 in 100, which is far from the risk level of 1 in 100,000 used by TCEQ.

TCEQ's own protocol for developing toxicity factors provided in Section A1.1 of Appendix 1 of the DSD states that one "extrapolate[s] from the adjusted POD to lower exposures based on MOA analysis"; however, as discussed above, in this DSD, modeling was done all the way down to a risk level of 1 in 100,000 using a (non-fitting) sublinear model. The DSD's approach is inconsistent with the guidance of EPA and other agencies (including possibly TCEQ as well, according to their protocol), in which a POD is selected near the low end of the observable range and then the mutagenic MOA established for EtO would support linear low-dose extrapolation.

f) See also Section A.2.20 of Appendix A of EPA's assessment (2016b) for more discussion of the above issues related to the lymphoid cancer risk estimates. The section critiques the approach used by Valdez-Flores et al. (2010), which was largely adopted in this DSD.

All of these scientific flaws contribute to the DSD's final unit risk estimate being a gross underestimate of the cancer risks demonstrated by the evidence.

Response:

a) See TCEQ's response to Comment #1.

b) See TCEQ's response to Comment #8.

c) The use of a 70-year cut-off in the lifetable analysis (or any other cut-off) is unrelated to the lifespan of the population. The purpose of the lifetable analysis is to calculate the extra risk of dying of the specific cause by age 70 years (irrespective of what the average lifetime is). Thus, if the goal was to calculate extra cancer mortality risk within the first 50 years of life, then the lifetable would calculate that extra risk by running the table up to 50 years. However, that does not mean that a 50-year old has a life expectancy of less than 50 years, or that he/she may not die from the specific cause after 50 years. Truncating the lifetable analysis at a specific age (say 70 years) does not mean that the risk of dying from the specific cause beyond the specific age is zero but that the *extra* risk of dying from the specific cause beyond that age is zero. The calculated extra risk is

due to the assumed exposure to the agent purportedly increasing the specific cause of death hazard rate.

USEPA (2016) acknowledges that by truncating their lifetable at age 85 years "EPA did not assume an 85-year lifetime" (page H-35 of USEPA 2016 appendices). The reason USEPA stopped at 85 is not because stopping at 85 years implies an average lifespan of 75 years. USEPA stopped the lifetable analysis at 85 years "because cause-specific disease rates are less stable for those ages" (footnote 16 on page 4-9 of EPA 2016 risk assessment) above 85 years. However, USEPA (2016) did not consider the fact that less than 1% of the workers in the NIOSH study lived past 85 years, which implies that the rate ratio models derived from the NIOSH study are not reliably applicable (stable) for these ages. In contrast, more than 15% of workers in the NIOSH study lived more than 70 years. Using proportional hazards models based on the NIOSH study to calculate extra risks beyond 70 years (e.g., 85 years) is extrapolating beyond the observed data used to derive the models.

The TCEQ decided to use 70 years to calculate the extra risk, based both on the science described above as well as the TCEQ policy to calculate the risk up to age 70 per TCEQ guidelines (TCEQ 2015). TCEQ policy takes into consideration the uncertainties associated with models derived from studies with short lifespan. TCEQ guidelines indicate the following [*emphasis added*]:

> If the probability of the specified response in the exposure-response model includes time (age), then the excess risk and the definition of the [effect concentration] EC also includes a specified time (age). For example, most exposure-response models used for epidemiology data incorporate the time at which a response is observed. In these models, the excess risk and the definition of the EC also include a specified time (age). In the calculation of excess risk, it is assumed that the exposure scenario remains unchanged up to that time (age) (i.e., a constant exposure concentration up to that age is presupposed). Furthermore, it is assumed that the estimated exposure-response model is appropriate up to that time (age). *For example, if an exposure-response model is estimated using occupational epidemiology data that only includes workers up to age 65 years, then calculating an excess risk up to age 70 years or higher involves an extrapolation over age that may or may not be warranted*. Also, because the excess risks and ECs are often heavily dependent upon the specified time (age), it is important to consider what the specified time (age) is when interpreting the results. For example, time might be age and the specified time be set to 70 years. In which case, the excess risk refers to the excess risk by age 70 years. *As is common in regulatory risk assessment, the TCEQ uses a default exposure duration of 70 years* as discussed in Chapter 1. *Another reason* to use an exposure duration of 70 years for calculation of excess risk using epidemiology data is that the *background rates of the disease and survival rates for a*

population used in the life-table analysis (BIER IV approach) discussed in the following section are more uncertain after 70 years.

d) TCEQ risk estimates are based on lymphoid mortality because the NIOSH study made available only the mortality data. Scientifically, it is incorrect and inappropriate to use a model based on observed data (e.g., lymphoid mortality) to predict values for different data (e.g., lymphoid incidence). The exposure-response relationship for mortality could be significantly different than the exposure-response relationship for incidence (e.g., USEPA (2016) modeling results for breast cancer mortality versus incidence were very different). The discrepancies between the two for TCEQ's critical cancer endpoint, lymphoid cancer, would depend on several factors including the survival period after detection of the specific cause of death.

   TCEQ is committed to using scientifically defensible approaches in its regulatory practices and strives to maintain both scientific accuracy and an appropriate level of conservatism (e.g., use of the 95% UCL, ADAFs). The TCEQ URF adequately protects human health because of the following conservative considerations:

   1) The lack of human data sufficient to establish EtO as a human carcinogen despite occupational exposures up to millions of times ambient levels;
   2) The lack of statistically significantly increasing risk with exposure to EtO;
   3) Using the endpoint (lymphoid cancer), study cohort (NIOSH), and sub-cohort (males) that predicts the largest estimated risk of seventy-two non-increasing exposure-response models;
   4) Uncertainty in the model parameter estimate (i.e., use of the 95% UCL);
   5) *Continuous* exposure to EtO, 24 hours a day, 7 days a week, for every week of every year, from birth until 70 years of age; and
   6) Adjustment for early-age exposures using ADAFs.

   Adding another layer of protection to the already protective evaluation would have to be a policy decision (i.e., weighed against a procedure resulting in technical/scientific inaccuracy).

e) The TCEQ evaluated the LEC at an extra risk of 1 in a 100,000 consistent with USEPA cancer guidelines (2005) on the selection of a POD at the low-end of the observable range of exposures. Although for animal studies, a typical POD is an extra risk of 0.10 because it corresponds to doses near the low-end of the doses, in epidemiological studies a lower level of risk needs to be used.

   TCEQ used the standard Cox proportional hazards model to calculate LEC for an extra risk of 1 in a 100,000 because the EC corresponding to the same risk level are in the range of the observed data in the NIOSH study. That is, the EC for an extra risk of 1 in 100,000 of lymphoid cancer mortality in males is 9.67E-03 ppm for 70 years with an exposure lag of 15 years, which correspond to a cumulative

occupational exposure of 591 ppm-days. There are 7 male workers in the NIOSH cohort with cumulative exposures less than 591 ppm-days. That is, 25.9% of the male workers in the NIOSH cohort that died with lymphoid cancer were exposed to cumulative exposures of less than the EC for 1 in a 100,000. In contrast, the EC for 1 in 100 results in environmental concentrations corresponding to cumulative occupational exposures of 354,400 ppm-days, which exceeds the largest cumulative exposure of lymphoid male decedents in the NIOSH study.

Table 1 above (see response to ACC Comment #15) shows the EC corresponding to different risk levels and the corresponding cumulative exposures with the number of lymphoid mortality cases of the male workers in the NIOSH study. The results in Table 1 show that the EC for an extra risk of 1 in a 100 is outside the range of cumulative exposures for the male lymphoid mortalities observed in the NIOSH study and in the upper 1% of cumulative exposures for all male workers. That is, all males that died with lymphoid cancers and more than 99% of all male workers had cumulative exposures less than EC(1/100). Thus, the NIOSH study does not support an extra risk of 1 in a 100 as a point of departure.

The EC for an extra risk of 1 in a 1,000 is a concentration that is in the high-end of cumulative exposures of male lymphoid mortalities observed in the NIOSH study. That is, 77.78% of all males that died with lymphoid cancers and 94.48% of all male workers had cumulative exposures less than the EC(1/1,000). Thus, a point of departure of 1 in 1,000 is at the higher-end of the cumulative exposures of male workers of the NIOSH study.

The EC for an extra risk of 1 in 10,000 is a concentration that includes 48.15% of the decedent men with lymphoid cancer and 66.45% of all men in the NIOSH cohort with smaller cumulative exposures. *The EC for an extra risk of 1 in 100,000 includes 25.93% of male lymphoid decedents and 30.17% of all males in the NIOSH study with smaller cumulative exposures.* Thus, use of an extra risk of 1 in 100,000 is supported by the NIOSH observed data, being near the lower end of the observed range of cumulative exposures to EtO, and is consistent with USEPA and TCEQ guidelines (USEPA 2005a, TCEQ 2015) on the selection of a POD at the low-end of the observable range of exposures.

Based on Table 1 results, using either 1 in 10,000 or 1 in 100,000 extra risk PODs (as PODs in the range of the observed data and close to the low-end of the observable range) round to the same ADAF-unadjusted URF selected by the TCEQ (2.5E-06 per ppb). Looking at it from a different perspective, using the 1 in 10,000 excess risk LEC of 4.04E-02 ppm as the POD and linear extrapolation, the 1 in 100,000 air concentration (ADAF unadjusted) is still 4 ppb (i.e., 1E-05/2.47E-06 per ppb = 4.05 ppb). This information has been added to Appendix 7 of the revised DSD.

f) Given the overall lack of evidence of the relationship between cancer risk and exposure to EtO (e.g., with correctly calculated p-values, models do not fit the lymphoid data statistically significantly better than the null model with zero slope), any unit risk estimate may overestimate the true cancer risk due to exposure to EtO, and therefore, be health-protective. The lower URF derived by TCEQ, compared to the USEPA URF, is primarily due to differences in the selected model. The TCEQ selected a model based on scientific considerations and principles that are discussed at length in the response to other comments.

Section A.2.20 of Appendix A of USEPA's 2016 assessment critiques the approach used by Valdez-Flores et al. (2010). The TCEQ has carefully considered those critiques and has made modifications to our approach based on their scientific validity. For example, the TCEQ has revised its assessment to incorporate ADAFs consistent with equation 5-17 of the TCEQ guidelines as opposed to incorporating them into the lifetable analysis as implemented by Sielken and Valdez-Flores (2009b) (see p. A-35 of Section A.2.20). The TCEQ did not agree with all of the USEPA's critiques however. For example:

> Because the Valdez-Flores et al. (2010) categorical results are for unlagged analyses, however, their referent groups are different from those used by Steenland et al. (2004). Valdez-Flores et al. (2010) used the lowest exposure quintile (providing there were sufficient data) as the referent group, whereas Steenland et al. (2004) used the no-exposure (lagged-out) group as the referent. [pg A-32]

This statement does not accurately represent how the Cox proportional hazards model estimates non-parametric rate ratios. In the Cox proportional hazards model the "referent group" is not an "unexposed" group. The "referent group" is the group of individuals alive at the time of the case being evaluated. That is, regardless of what the exposure lag is, the "referent group" is the same for each individual at that particular timepoint. The rate ratios may change between unlagged and lagged cumulative exposures because the cumulative exposure (which is the main covariate) changes (this is true for Steenland et al. (2004) and Valdez-Flores et al. (2010) analyses). Statements like this in USEPA's risk assessment have resulted in many readers misunderstanding results of exposure-response modeling. In fact, it was in response to Valdez-Flores et al. (2010, 2013) publications through SAB recommendations that USEPA changed their dose-response modeling of a linear model using quintiles in 2006, 2011, and 2013 to modeling the individual data in 2016. The TCEQ also uses the individual data, and among other important considerations, demonstrates that: (1) USEPA's two-piece spline model does not fit the data better than the Cox model; and (2) USEPA's selected model assessment statistically overestimate the number of lymphoid cancers in the NIOSH cohort as a whole and for the individual exposure quintiles whereas TCEQ's model is relatively accurate. The TCEQ DSD extensively documents and justifies use of the Cox proportional hazards modelling results based on these and other scientific considerations.

Comment 12:

The DSD does not appropriately account for the science showing increased cancer risks from early life exposures to carcinogens with a mutagenic mode of action.

The DSD states that the approach of Sielken and Valdez-Flores (2009) was used to apply the age-dependent adjustment factors (ADAFs) to the cancer risk estimates; however, the ADAF calculations were not done correctly by Sielken and Valdez-Flores (2009). Early life exposures to chemicals with a mutagenic MOA such as EtO can increase lifetime cancer risk, and thus EPA guidance recommends the application of ADAFs in quantitative risk calculations to adjust for this potential increased susceptibility. This means that exposure to a mutagenic carcinogen at a young age can increase a person's risk of developing cancer later in life. Thus, the ADAFs are designed to adjust lifetime risk, to reflect increased lifetime cancer risk from increased susceptibility to early-life exposures. But Sielken and Valdez-Flores (2009) incorrectly multiply the ADAFs to the age-specific cancer mortality rates in the lifetable, which just applies the factors to risk for those younger age groups and ignores increased risks for older ages (discussed in more detail in EPA's assessment). In addition, assuming increased early-life susceptibility and applying the ADAFs along with the Cox PH model in the lifetable analysis, as done by Sielken and Valdez-Flores (2009), is inconsistent with a major assumption of the Cox model, that RR is independent of age.

In fact, because of the lagged exposures and low cancer mortality rates at young ages, applying the ADAFs just to young age groups had a negligible effect on the final risk estimates in Sielken and Valdez-Flores (2009). In contrast, the approach that correctly accounts for the science showing that early life exposures increase lifetime cancer risks (used by EPA) increased the lifetime risk estimates by about 22% (for both female breast cancer and lymphoid cancer combined).

The DSD's confidence score for sensitive populations was "medium." However this score is not warranted because TCEQ discounts the breast cancer risk in females and misapplies the ADAFs for susceptibility from early-life exposures, both of which result in underestimations of the risks posed by ethylene oxide.

Response:

The TCEQ DSD now calculates the ADAF-adjusted long-term ESL consistent with equation 5-17 of the TCEQ guidelines (TCEQ 2015), rounding it to two significant figures at 2.4 ppb. That being said, we do want to address some of the points made concerning the implementation of ADAFs by Sielken and Valdez-Flores (2009).

The Sielken and Valdez-Flores (2009) paper outlines issues regarding the use of ADAFs developed for animal studies that are exposed to a constant dose throughout their life. The article indicates that ADAFs as developed were meant for models that assume a constant dose metric (as is the case for most animal-based dose response models) and not for age-dependent dose metrics (as is the case for the models developed for EtO based on the NIOSH study). Sielken and Valdez-Flores (2009) work through a USEPA example to show that the ADAFs as

developed by USEPA do not apply to a model that uses cumulative exposure as a dose metric, as opposed to a model that uses average exposure as a dose metric.

TCEQ guidelines use ADAF for models based on a constant dose metric using USEPA (2005) guidelines. However, for models based on cumulative dose metrics (the dose metric used for EtO), the TCEQ guidelines allow for implementation of ADAFs in lifetable analyses (see Section 7.9.4 of TCEQ 2015). However, in the present case the TCEQ has elected to revise its assessment to incorporate ADAFs consistent with equation 5-17 of the TCEQ guidelines as opposed to incorporating them into the lifetable analysis since USEPA (2016) specifically objects to the ADAF implementation discussed in Sielken and Valdez-Flores (2009) (see p. A-35 of Section A.2.20 of USEPA 2016).

Comment 13:

The DSD uses a scientifically inappropriate comparison explicitly rejected by the SAB to "predict" the numbers of cases in the NIOSH cohort.

a) TCEQ's method for predicting the number of cases in the NIOSH cohort relies on a standardized mortality ratio (SMR) comparison. As discussed in Comment #10b, SMRs are notoriously deficient for analyzing occupational epidemiology data because workers often have background disease mortality rates below those of the general population. This concept is called the "healthy worker effect" (HWE), although it can reflect differences between an occupational cohort and the general population beyond health. In fact, EPA's SAB specifically recommended that epidemiological results based on external standards, e.g., SMRs, be down-weighted. The SAB states "The presence of the healthy worker effect cannot be denied in these occupational data and the use of an external standard for comparison does not avoid healthy worker types of biases."

   In the DSD, basing the "predictions" on an SMR comparison ignores the healthy worker effect apparent in the data and inflates the background risk expected in the cohort, equating it to the background risk in the general population. Therefore, all the relative risk (RR) models, which are based on an internal analysis estimating increases in risk relative to the actual (lower) background rates in the cohort, will overestimate cohort case numbers when the increases in risk are forced to be relative to the higher background rates of the general population. This will be true unless they're underestimating the risks to begin with, like the sublinear model selected by the TCEQ. The selected EPA models naturally "overpredict" case numbers under this flawed approach.

b) Instead, if one performs a more appropriate comparison based on the results of internal analyses (within the cohort), one can see that the sublinear model used by TCEQ is a poor predictor of the nonparametric categorical RR estimates for the exposure quartiles; see model "e^(β*exp)" in Fig 4-3 of EPA's assessment.

Comparing TCEQ's model, depicted by the solid blue curve near the bottom of the graph, to the nonparametric categorical RR estimates, depicted by the filled purple circles, shows that the model selected by the TCEQ substantially underestimates the nonparametric categorical RR estimates. In contrast, the EPA model depicted by the dashed red line (linspline1600) is a much better predictor of the nonparametric categorical RR estimates. As noted in Comment #8, the nonparametric baseline estimate is the best available estimate of the baseline hazard rate in the cohort because it is based on unexposed referent group without any assumptions about the shape of the exposure-response model for the exposed workers and, thus, without any influence of the higher-exposure data on the model fit to the lower-exposure data. Similarly, the categorical RR estimates for the exposed groups are estimated with no assumptions about the shape of the exposure-response relationship across the groups.

In addition, proper comparisons of models against data should be based on maximum likelihood estimates (MLEs), as done in Figure 4-3 of EPA's assessment, not upper bounds as primarily reported by TCEQ.

Thus the DSD's reliance on this flawed calculation to support its rejection of EPA's model and its own use of a poorly fitting model is not supported by the evidence.

Response:

This response provides both a more general response and a much more detailed response complete with additional statistical analyses.

a) The premise of the comment is that a healthy worker effect should be assumed for the cancer endpoints being evaluated for EtO. However, as discussed in response to Comment #10, the TCEQ finds that robust data from Kirkeleit et al. (2013) indicate the lack of a healthy worker effect for both lymphoid/hematopoietic and breast cancer incidence. In addition, the SMRs for workers in the NIOSH and DOW/UCC are similar to those reported by Kirkeleit et al. (2013) for lymphoid cancer mortality and show no evidence of a healthy worker effect (see Section A3.3.1 of the DSD).

b) As discussed in response to Comment #6, categorical RRs should not be used for visually comparing models fit to individual data, particularly when appropriate statistical model fit criteria are available. More specifically, estimated nonparametric RRs are calculated with respect to an underlying background hazard rate that is also estimated nonparametrically. The RRs of parametric models fit to the individual data are defined with respect to an underlying background hazard rate estimated by the model. However, the underlying background hazard rates estimated by the nonparametric RRs and the parametric model are generally different. A better comparison of models fit to the observed data is to use the predictiveness of the model; that is, the capability of the model to estimate the observed number of deaths with a certain degree of confidence (see Appendix 2 of the revised DSD). Moreover, visual interpretation of the consistency of categorical RRs with the shape/slope of a modelled dose-response can change as the

number of exposure categories changes (e.g., Figures 1-3 and supplementary materials of Valdez-Flores and Sielken 2013). The visual presentation of only a few exposure categories can blind the data user to the variability in the underlying dose-response data, and by corollary, preclude an appropriate visual assessment/comparison of model fit to the actual individual data. See Appendix 5 of the revised DSD for additional information.

In consideration of the comment that proper comparisons of models against data should be based on maximum likelihood estimates (MLEs), MLEs have been added to Figures 9-12 of the DSD.

The following provides more detailed responses to comments a) and b).

Detailed Response to Comment a)

The approach used by TCEQ to predict the number of cause-specific deaths in the NIOSH cohort use the SMR methodology to evaluate the models. The approach for calculating SMRs is well established and has been used by regulatory agencies and researchers to compare mortality rates in target populations to mortality rates in reference populations. Thus, the approach used is well documented and has been extensively used.

The models used by TCEQ were derived using internal comparisons and did not rely on the general U.S. population background mortality rates. U.S. population background rates are used to predict risks in the general population. The SAB comment about the healthy worker effect (HWE) was considered in the evaluation predictions reported by TCEQ. The HWE has been shown to be cause-specific in incidence studies. That is, there are endpoints for which workers experience a HWE and some endpoints for which the HWE is negative, and yet other endpoints for which the HWE is non-existent.

Kirkeleit et al. (2013) researched the HWE in 366,114 randomly selected Norwegian workers and compared the incidence of several endpoints with the general Norwegian population. Their findings indicate that there is potential for the HWE for some endpoints and increased incidence (i.e., an "unhealthy" worker effect) for other endpoints. *For lymphoid and hematopoietic cancer incidence Kirkeleit et al. (2013) did not find a HWE* with SIRs and 95% confidence intervals of 0.97 (0.90, 1.03) and 1.09 (0.92, 1.27) for male and female workers, respectively. That finding was also true for breast cancer with an SIR and 95% confidence interval of 1.02 (0.95, 1.09).

SMRs for workers in the NIOSH and DOW/UCC are similar to the SIRs reported by Kirkeleit et al. (2013) for lymphoid cancer mortality. That is, the mortality rate in unexposed male and female workers in the NIOSH study are not statistically significantly different from the mortality rate of the general U.S. population. Again, footnote * to Table 32 in TCEQ DSD states "Quintile 1 is the control (unexposed lagged-out) group with 9 lymphoid cancer mortalities observed and 11.5 mortalities predicted by all models with a 95% confidence interval of (6.0 and 25.2), which

includes the 9 lymphoid cancer deaths." This confidence interval indicates that the observed 9 lymphoid cancer deaths in the unexposed male and female workers of the NIOSH cohort is consistent with the number of lymphoid cancer deaths in the general U.S. population during the same period of time after accounting for age, sex, and calendar year. Expressed in terms of SMRs, the SMR for lymphoid cancer deaths in the unexposed male and female NIOSH workers is equal to 0.78 (=9/11.5) with a 95% confidence interval (CI) equal to (0.36, 1.50). The 95% CI on the SMR for unexposed workers includes the value of one, which indicates that *the mortality rate in the unexposed workers in the NIOSH study and the U.S. population mortality rate are not statistically significantly different* at the 5% significance level.

Kirkeleit et al. (2013) indicates that there is minimal, if any, HWE for mortality from cancer. Kirkeleit et al. (2013) estimates an overall cancer SMR of 0.85 and 0.84 for male and female workers, respectively. Despite: (1) that the mortality rate in the unexposed workers in the NIOSH study and the U.S. population mortality rate are not statistically significantly different (as discussed above); and (2) the lack of a HWE for more EtO-relevant lymphoid and hematopoietic cancer based on data from Kirkeleit et al. (as discussed above), the TCEQ conducted a sensitivity analysis assuming that these overall cancer SMRs apply to lymphoid cancers. That is, despite data to the contrary, the TCEQ sensitivity analysis assumes NIOSH workers had lesser cancer mortality than the general population by multiplying the U.S. male and female background hazard rates by 0.85 and 0.84, respectively, to account for the assumed HWE. The results did not change significantly. *USEPA's selected model still statistically significantly overestimates the number of observed (53) lymphoid deaths in the NIOSH study; 77.5 with a 95% CI of (59.3, 103.6)*. By contrast, the standard Cox proportional hazards model includes the observed number (53) of lymphoid deaths in the NIOSH study within the 95% confidence interval; 44.3 with a 95% CI of (33.9, 59.2). *Thus, even using this conservative HWE (in the face of more study-specific data to the contrary), USEPA's selected model significantly overestimates the observed data*.

<u>Detailed Response to Comment b)</u>

Part b) of Comment #13 is the same issue raised in Comment #6. Since these comments have been repeated, the response that follows is similar to the response for Comment #6.

There is over-reliance on Figure 4-3 of USEPA's risk assessment to judge model fit to lymphoid mortality in the NIOSH epidemiological study. This over-reliance and misinterpretation is despite the note USEPA wrote in the Figure 4-3 that 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.)" The misinterpretation of model fitting versus non-parametric rate ratios was discussed at length by Valdez-Flores et al. (2013), and USEPA (2016) responded by adding the note quoted above to the figures showing rate ratios. However, USEPA (2016) still includes statements like "adequate visual fit" in several places of their risk assessment. There are several reasons why this Figure 4-3 and other similar figures showing rate ratios from different models

could be misleading to casual readers and these are described in detail in response to Comment #6.

Comments from the Union of Concerned Scientists

Comment 1:

The TCEQ's proposed cancer risk value is not an appropriate replacement for the EPA IRIS risk value. The EPA IRIS assessment on the carcinogenicity of ethylene oxide issued in 2016 incorporated the best available science, public comment opportunities, interagency review, and scientific peer review by EPA's Scientific Advisory Board.

Response:

The TCEQ is not proposing to replace the IRIS value. The TCEQ is deriving the agency's own value, and in doing so, evaluated the USEPA's 2016 EtO derivation. As documented in the DSD, the TCEQ determined that the USEPA's EtO URF lacked scientific justification and did not pass reality checks on lymphoid cancer predictions for the key worker cohort and the U.S. general population, as well as calculating incorrect model fit criteria. The TCEQ dose-response assessment has already undergone internal peer review and public comment and will also undergo independent external expert peer review in the first quarter of 2020.

Comment 2:

The proposed TCEQ risk value uses a sublinear dose response model to derive its toxicity factors that the EPA would deem inappropriate due to ethylene oxide's reactive, mutagenic, and multisite carcinogenicity.

Response:

The Cox proportional hazards model is a standard dose-response model and is commonly used by TCEQ, USEPA, and other regulatory agencies for dose-response assessment. Although described as sublinear, it is indistinguishable from linear across all doses of interest (as shown in many figures in the DSD such as Figure 22) and is used by TCEQ in calculating a point of departure for linear low-dose extrapolation. USEPA considered results from the Cox proportional hazards model as it is a model USEPA deems appropriate for chemicals such as EtO, although USEPA did not select it based on an assessment of model fit that TCEQ does not agree with (e.g., USEPA incorrectly calculated p-values and AIC values). While a model that is indistinguishable from linear across doses of interest is appropriate considering EtO's carcinogenic MOA (as is linear low-dose extrapolation), the TCEQ notes that USEPA acknowledges it cannot provide mechanistic support for its overall supra-linear model.

Comment 3:

Further, the TCEQ assessment relies on a key study that EPA chose not to include in its evaluation of the best available science because the data "were not of sufficient quality" and other recent studies considered in the TCEQ's assessment would benefit from further conflicts of interest scrutiny as several are funded by the American Chemistry Council, which has a vested interest in ethylene oxide production and regulation.

Response:

The TCEQ does not rely on the referenced UCC cohort study to derive the EtO URF. While the UCC cohort data was evaluated, the TCEQ assigned 100% of the URF weight to the same NIOSH cohort as USEPA (2016) relied on.

Science is paid for and conducted by those who have interest in better understanding the underlying science and important scientific issues, which can include government, industry, academia, and others. The TCEQ used a systematic review process that included a risk of bias analysis for all articles (including the published peer-reviewed scientific articles being referenced), which is fully described in Appendix 1 of the DSD.

Comment 4:

Choosing to abandon EPA's IRIS risk value for ethylene oxide, especially without an external review process, would mean a departure from the use of best available science.

Response:

The TCEQ thoroughly evaluated the USEPA 2016 IRIS EtO URF and concluded that it did not represent the best-available science. The TCEQ's dose-response assessment has undergone internal peer review and public comment and will also undergo independent external expert peer review in the first quarter of 2020.

Comment 5:

Data on ethylene oxide released by EPA's National Air Toxics Assessment (NATA) in 2018 revealed that the chemical is significantly contributing to higher cancer rates in areas surrounding chemical manufacturers and sterilizers using the chemical across the country. About a quarter of the facilities contributing to these risks are located in Texas. According to the NATA data, the probability of developing cancer from air pollutants was beyond the EPA's acceptable level of risk in over 100 communities, and 91 percent of the risk can be attributed to ethylene oxide, formaldehyde, or chloroprene.

Response:

USEPA's NATA produces theoretical risk estimates based on emissions inventories and IRIS dose-response assessments. Because there are serious concerns about the USEPA's EtO IRIS assessment (discussed in detail in the DSD and in response to public comments in this document), then so too is the NATA evaluation questionable. The NATA has not revealed that EtO is significantly contributing to higher cancer rates in areas surrounding chemical manufacturers and sterilizers, nor has the NATA demonstrated that those areas actually have higher cancer rates.

Comment 6:

Further, EPA issued its findings from air monitoring outside of the Sterigenics facility in Willowbrook, IL that was shut down by the state, comparing emissions before and after the shutdown. The monitors revealed levels 90 percent lower at the sites closest to Sterigenics,

illustrating the direct relationship between the facility's operations and ethylene oxide levels. A March report from the Illinois Department of Health found that cases of Hodgkin's lymphoma among women in the Willowbrook community were nearly 90 percent higher than in a nearby county. This data is in agreement with the systematic review conducted by IRIS that evaluated the toxicological and epidemiological evidence available on the chemical and determined that it was carcinogenic to humans, and exposure was linked to an increased risk of cancer of leukemia, lymphoma and breast cancer in women.

Response:

Hodgkin's lymphoma is not a cancer included in USEPA's dose-response assessment; non-Hodgkin's lymphoma is. Regardless, such ecological studies are not used and are insufficient for dose-response assessment, having much more uncertainty (particularly for the exposure assessment) and much lower numbers than the 17,000+ worker study that both USEPA and TCEQ relied on. The referenced assessment by the Illinois Department of Health indicated [*emphasis added*]:

> Living in a study area at the time of diagnosis was used to represent potential exposure to EtO, but it was a very crude proxy…data on actual exposure in individuals was non-existent…there is considerable uncertainty about the length and the level of exposure to EtO that each individual in Willowbrook, and surrounding areas, may have actually experienced in the past. *Any observed increase, in and of itself, is insufficient to draw conclusions regarding the potential impact of EtO exposure*.

In contrast to the Hodgkin's lymphoma findings referred to by the commenter, the following are findings from similar studies around EtO emitters [*emphasis added*]:

(1) The Colorado Department of Public Health and Environment found no evidence that there is more cancer in the communities around Terumo BCT than in surrounding areas (https://drive.google.com/file/d/1WEe0kCfkXW2RQC4jRFsIC803u_6P1Mub/view). Their analysis examined all cancers combined, and five individual types of cancer: Hodgkin's lymphoma, non-Hodgkin's lymphoma, multiple myeloma, lymphocytic leukemia, and breast cancer (females only). They state, "*The incidence of all cancers combined and five individual types of cancer in the community surrounding Terumo BCT were no different than expected* based on cancer rates in the remainder of Colorado for the years 2000 through 2017."

(2) The Michigan Department of Health and Human Services evaluated cancer incidence in the area surrounding Viant Medical Facility in Grand Rapids in response to the discovery of elevated levels of EtO in the area (https://www.michigan.gov/documents/mdhhs/Viant_Cancer_Incidence_Review_661354_7.pdf). Cancer registry incidence data was evaluated for female breast cancer, multiple myeloma, Hodgkin's lymphoma, non-Hodgkin's lymphoma, and leukemia. No statistical elevations in the frequency of any of the cancer types were observed over the 15-year time period of analysis, with the exception of multiple myeloma when compared to county rates but not state rates (see the table below; small number of cases resulting in an imprecise estimate with borderline statistical significance). The Michigan Department of Health and Human Services concludes

[*emphasis added*], "*Therefore, the results of the analyses presented in this report do not suggest that further investigation is needed at this time*."

Table 2: Standardized Incidence Ratio (SIR)[1] for Invasive Cancers by Cancer Type, Comparing the Geographic Area of Analysis (10 Census Tracts) to Kent County and the State of Michigan, Adjusted for Age and Sex, 2001 – 2015[2]

		Kent County Comparison		State of Michigan Comparison	
Cancer Type	Observed Number of Cases in 10-Census Tract Area	Expected Number[1] of Cases	SIR (95% Confidence interval)	Expected Number[1] of Cases	SIR (95% Confidence interval)
Female Breast	177	219.48	0.81 (0.71, 0.91)	202.23	0.88 (0.77, 0.99)
Non-Hodgkin Lymphoma	71	66.76	1.06 (0.86, 1.30)	68.19	1.04 (0.85, 1.27)
Hodgkin Lymphoma	12	14.85	0.81 (0.47, 1.31)	15.07	0.80 (0.46, 1.29)
Multiple Myeloma	25	16.96	1.47 (1.02, 2.06)	19.28	1.30 (0.90, 1.81)
Leukemia	43	44.91	0.96 (0.73, 1.23)	48.06	0.89 (0.68, 1.15)

[1]SIR is the ratio of observed to expected cases, where expected cases are calculated by multiplying the age-sex specific cancer incidence rates for the comparison population (Kent County or State of Michigan) to population estimates for the 10-census tract region surrounding Viant (2010 US Census).
[2]Source of Michigan Cancer Cases: Michigan Resident Cancer Incidence File, 2001-2015, includes cases processed through November 30, 2018. Division for Vital Records & Health Statistics, Michigan Department of Health and Human Services. Cancer cases were defined by new malignant cancer diagnoses where the listed cancer site was the primary site.
Shading indicates that the SIR was statistically higher or lower than 1.00 (95% confidence interval does not include the value 1.0)

As can be seen in the table above, Hodgkin's lymphoma incidence is somewhat less than expected (in contrast to Willowbrook, IL), and breast cancer (one of the two endpoints used by USEPA 2016) is actually statistically significantly decreased around this EtO-emitting facility, using both county and state rates as comparators. Therefore, assessing evidence in a balanced manner demonstrates that there are mixed and largely insignificant associations between EtO and cancer in these communities, and that any single such ecological study cannot be cited as convincing evidence of a chemical's carcinogenicity (or lack thereof).

Comments from the Environmental Defense Fund (EDF)

Comment 1:

TCEQ should not reduce a risk threshold to EtO, a known human carcinogen. The public health issues related to EtO pollution recently attracted national attention when Willowbrook, Illinois began to experience extremely high rates of lymphatic cancers which was directly linked to high concentrations of EtO emitted into the air from a medical equipment sterilization plant. Fully aware of these risks, TCEQ now proposes to move in the opposite direction by raising the threshold for EtO exposure to 4 ppb, a massive jump from the current air permit screening level of 1 ppb. This change is further significant because of the volume of EtO Texans are already exposed to under the current limit – Texas is responsible for around half of the EtO emissions in the United States. TCEQ should abandon its flawed approach in its proposed assessment and adhere to its clear mandate to protect public health and welfare.

Response:

As discussed in the response to the Union of Concerned Scientists comments, a full assessment of ecological epidemiology studies demonstrates that there are mixed and largely insignificant associations between EtO and cancer in these communities. Although there are significant uncertainties involved, the TCEQ provides the following for additional context.

Based on the 2014 National Emissions Inventory (NEI), Texas emits approximately 36% of the EtO in the US. As a result, tons of EtO emitted per square mile in Texas (1.8E-04 tons/square mile) is over 5 times higher in Texas compared to the rest of the US (3.5E-05 tons/square mile). Despite this and the very high carcinogenic potency purported by USEPA (2016) for EtO causing lymphoid and breast cancers, the incidences of leukemia and non-Hodgkin's lymphoma (both included in USEPA's dose-response assessment) as well as breast cancer are *lower* in Texas than in the general US population, with the same being true for all cancers combined (Table 3). Again, leukemia, non-Hodgkin's lymphoma, and breast cancer are endpoints included in USEPA's carcinogenic dose-response assessment for EtO (USEPA 2016), along with multiple myeloma (for which state-specific versus US data were not available).

Table 3. Relevant Age-Adjusted Cancer Incidence Rates per 100,000 (2012-1016)

Area	NEI Emissions (tons)	Emissions per Area (tons/mile^2)	Non-Hodgkin's Lymphoma	Leukemia	Breast Cancer (female)	All Cancer Sites
US	133.72	3.52E-05	19.2 (19.1, 19.3)	14.1 (14.1, 14.2)	125.2 (124.9, 125.4)	448.0 (447.7, 448.4)
Texas	48.45	1.80E-04	17.4 (17.2, 17.6)	13.9 (13.7, 14.1)	111.9 (111.2, 112.7)	407.7 (406.6, 408.9)
Jefferson County	12.05	1.08E-02	17.5 (15.3, 19.9)	13.4 (11.5, 15.5)	102.4 (94.9, 110.3)	399.9 (389.3, 410.7)
Harris County	11.75	6.60E-03	16.9 (16.3, 17.5)	13.0 (12.5, 13.5)	111.9 (109.9, 114.0)	400.1 (397.2, 403.1)

At the county level, although highly-industrialized Jefferson County (population ≈260,000) has more EtO emissions on a square mile basis than any other county in Texas (1.1E-02 tons/square mile) with over 300 times more than the US at large (3.5E-05 tons/square mile), the incidences of leukemia (13.4 (95% CI of 11.5, 15.5)), non-Hodgkin's lymphoma (17.5 (95% CI of 15.3, 19.9)), and breast cancer (102.4 (95% CI of 94.9, 110.3)) are lower in Jefferson County Texas than in the general US population (Table 3). In fact, breast cancer incidence is statistically significantly lower in Jefferson County compared to both Texas and the US, despite EtO emissions that are 60 times higher than Texas at large and 307 times higher than the US. Similarly, highly-industrialized Harris County is by far the most populated Texas county (≈4.6 million) with relatively high reported NEI EtO emissions per square mile (i.e., 6.6E-03 tons/square mile is ≈188 times higher than the US at 3.5E-05 tons/square mile), the incidences of leukemia (13.0 (95% CI of 12.5, 13.5)), non-Hodgkin's lymphoma (16.9 (95% CI of 16.3, 17.5)), breast cancer (111.9.4 (95% CI of 109.9, 114.0)), as well as all cancers combined (400.1 (95% CI of 397.2, 403.1)) are all statistically significantly lower than in the general US population (Table 3). Despite the associated uncertainties, these results are not consistent with EtO as a potent carcinogen causing observable increases in lymphoid or breast cancer at environmental concentrations.

Lastly, the sound scientific basis of the TCEQ assessment is set out in the DSD and is also addressed in other comment responses in this document. The DSD demonstrates that it is USEPA's 2016 assessment that is flawed, being over-predictive of lymphoid cancers in both the key cohort and the U.S. general population, using incorrectly calculated p-values and AIC values, and lacking MOA data to justify their non-standard overall supra-linear dose-response model, culminating in risk estimates shown to be biologically implausible.

Comment 2:

TCEQ's proposed assessment for EtO is flawed scientifically. The most egregious scientific error in TCEQ's proposed assessment is the dose-response analysis using the Cox proportional-hazards model fit to these data by Valdez-Flores et al. (2010) (Valdez-Flores et al. 2010). In the analysis, TCEQ estimates the slope of the dose response curve using the lowest exposure category rather than the 100 randomly matched unexposed individuals as done by Steenland et al. (2004) (Steenland et al. 2004). This is by far and away the most important difference between EPA's cancer risk analysis of ethylene oxide and the TCEQ analysis.

Response:

This comment may be due to a misunderstanding of how Cox proportional hazards models are fit to epidemiological data. The comment states that the Valdez-Flores et al. (2010) model fit uses the lowest exposure category rather than unexposed individuals. However, the Cox proportional hazards model does not use a control unexposed group to derive the slope of the model. Rather, Cox models define a risk set for every case (e.g., every cancer death from the specific cause) that includes all the individuals that are at risk at the time the case occurred (e.g., the time of the cancer mortality from the specific cause). The risk set includes exposed and unexposed workers (see e.g., Allison, 2010 and Steenland et al. 2004). Thus, every case in the NIOSH study has, possibly, more than 17,000 individuals in the risk set. The Valdez-Flores et

al. (2010) models and TCEQ models used the full risk set (which includes unexposed and exposed individuals) for every case.

USEPA's model fitting, developed by Professor Kyle Steenland, on the other hand, rather than using the full risk set with all individuals for every case, uses only 100 individuals randomly sampled from the full risk set. [EDF comments erroneously indicate that in Steenland et al. (2004), 100 individuals selected for the risk sets include only "unexposed individuals"; however, the "controls" in this analysis were actually exposed individuals from the dataset who had not yet died by the age of the cause-specific decedent.] Steenland used only 100 to speed up computer run time, possibly at the expense of precision. This is clearly stated in Steenland et al. (2004) where he writes, "risk sets were constructed in which 100 randomly selected controls were chosen for each case from the pool of all those who survived without haematopoietic or breast cancer to at least the age of the index case. Use of 100 controls has been shown to result in virtually the identical rate ratio with all possible controls (the full risk set), with approximately the same precision, while making possible more rapid computer runs." Although USEPA's and Steenland's approximation using only 100 randomly chosen individuals to form the risk set for every case may result in estimates that have "approximately the same precision", these estimates are not easily reproducible because of the randomness in the selection of the 100 individuals used instead of the full risk sets.

Based on the discussion above, it is clear that how the TCEQ and Dr. Steenland models were actually fit (i.e., full risk set versus 100 individuals randomly sampled from the full risk set) is not the most important difference between the USEPA and TCEQ analyses. Rather, the most important difference between USEPA's and TCEQ's assessments are the functional form of the exposure-response model. USEPA used a non-conventional exposure-response relationship that is supra-linear overall (two-piece linear model), whereas TCEQ used a standard exposure-response model that while sublinear in nature (standard Cox proportional hazards model) is indistinguishable from linear over the doses of interest in this case. Furthermore, TCEQ showed that USEPA's two-piece linear model did not fit the NIOSH data better than the standard Cox proportional hazards model (see Appendix 4 of the revised DSD).

Comment 3:

The obvious place to see the effect of this assumption is on page 133 of the TCEQ DSD document (Figure 22). It is clear that the dashed blue line used by TCEQ does not pass through 1 when the x-axis (15-year lagged ppm-days) equals zero, but instead almost passes through EPA's relative risk for the first quintile. This results in an over-estimation of the risk in the unexposed population quite dramatically, in fact treating this population as if the risk were the same as the risk in the first quintile. Because the relative risk climbs dramatically from the control response to the first quintile, using the first quintile as the denominator in the risk ratio for Cox modeling substantially reduces the slope, and thus the risk.

Surprisingly, TCEQ illustrates how bad their fit is relative to the control population in Figures 20 and 21 of the DSD. In these figures, rather than having the model run through the red dot for the first quintile, TCEQ's model (the blue solid line) is forced through the relative risk of 1 for

the unexposed group. We illustrate this point by placing the blue line (now a black thick line) running directly through the first quintile red dot on the same plot in Figure 21. One can see the resemblance to the lines shown in Figure 23 [sic].

By calculating the slope relative to the lowest exposure category (>0-1199 ppm days) instead of relative to unexposed, TCEQ has disregarded all of the increased risk at low external exposures in this data set and has calculated a slope factor that is meaningless. The arguments for the fit of the model are not convincing since TCEQ disregards the fact that their model is overestimating response in the unexposed group. Indeed, the fit of the TCEQ model as it is being used for risk calculations is the solid blue line passing through 1.0 in Figures 21 and 22. It is clear that the model used does not fit these data. Finally, TCEQ has ignored the fact that additivity to background is likely to lead to linear or supra-linear response.

Response:

As a note, the proposed DSD contained no Figure 23. That being said, the TCEQ figures are apparently being misinterpreted. *The rate ratios of TCEQ's model do go through 1* as estimated by the Cox proportional hazards model and shown in Figures 19, 20, and 21 (lowest blue solid line) of the proposed DSD. However, as USEPA recognized in similar graphs (one of them reproduced in Figure 19 of the proposed DSD) the models are not comparable on the y-axis. USEPA (2016) stated in their Figure 4-3 footnote [*emphasis added*], "Note that, with the exception of the categorical results and the linear regression of the categorical results, *the different models have different implicit estimated baseline risks; thus, they are not strictly comparable to each other in terms or RR values*, i.e., along the y-axis. They are, however, comparable in terms of general shape." The TCEQ sequence of graphs was meant to illustrate that the alleged bad fit of the standard Cox proportional hazards model to four rate ratios estimated for quintiles of grouped decedents is an artifact caused by the difference of the implicitly estimated baseline hazards. It is important to note first that rate ratios for the quintiles are nonparametric *estimates* of the rate ratios and are *not observed* rate ratios.

A much better representation of the spread, variability, and uncertainty surrounding the rate ratios is depicted in Figure 20 from the DSD (reproduced as Figure 9 below).

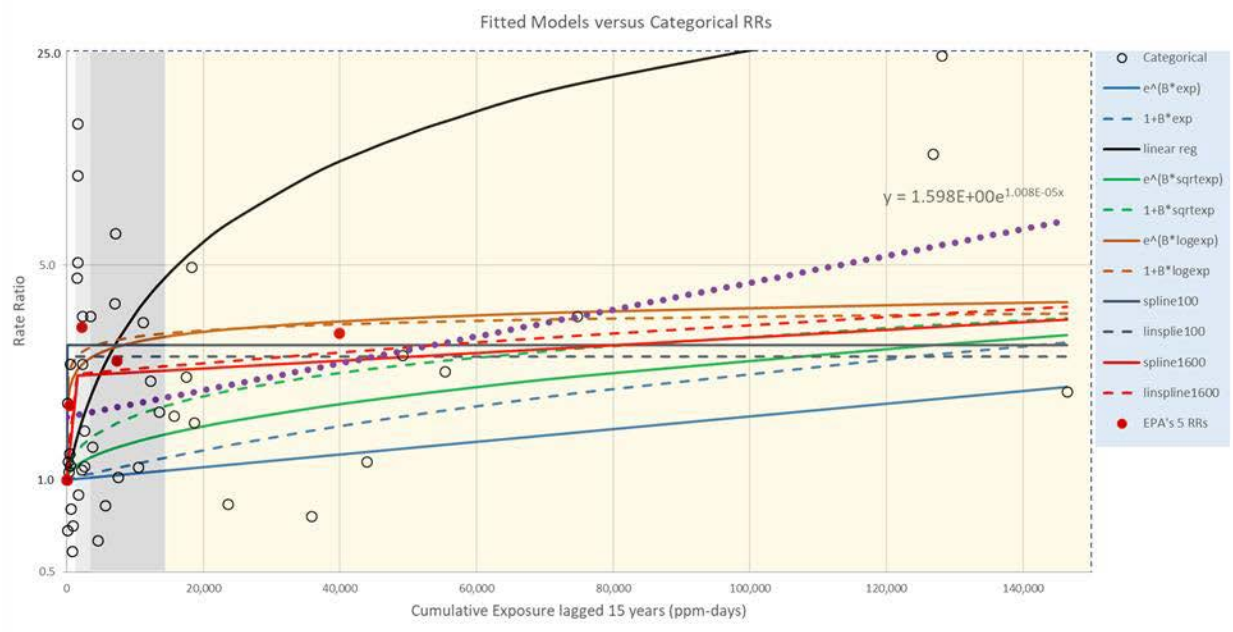


Figure 9. Lymphoid Cancer Death Categorical Rate Ratios (RRs) and Various Fitted Models for 15-Year Lagged Occupational Doses ≤150,000 ppm × days (NIOSH cohort)

Figure 9 above, which is Figure 20 of the DSD, shows both the estimated rate ratios for USEPA's quintiles and the estimated rate ratios for each individual case of lymphoid mortality. This figure shows the same models and data shown in Figure 19 of the DSD, but includes: (1) the estimates of the individual rate ratios (circles); (2) the y-scale in logarithmic increments for better resolution of the RRs; and (3) a line (purple-dotted line) fitted to the estimated risk ratios of the individual lymphoid decedents using least squares. The background colors in Figure 9 (Figure 20 of the DSD) show the range of lymphoid decedents included in each of USEPA's summary quintile RRs shown as red dots.

Figure 10 below, which is Figure 22 of the DSD, is similar to Figure 9 above (Figure 20 of the DSD) but is limited to showing only the two-piece linear spline model developed by USEPA and the log-linear model developed by TCEQ for visual comparison.

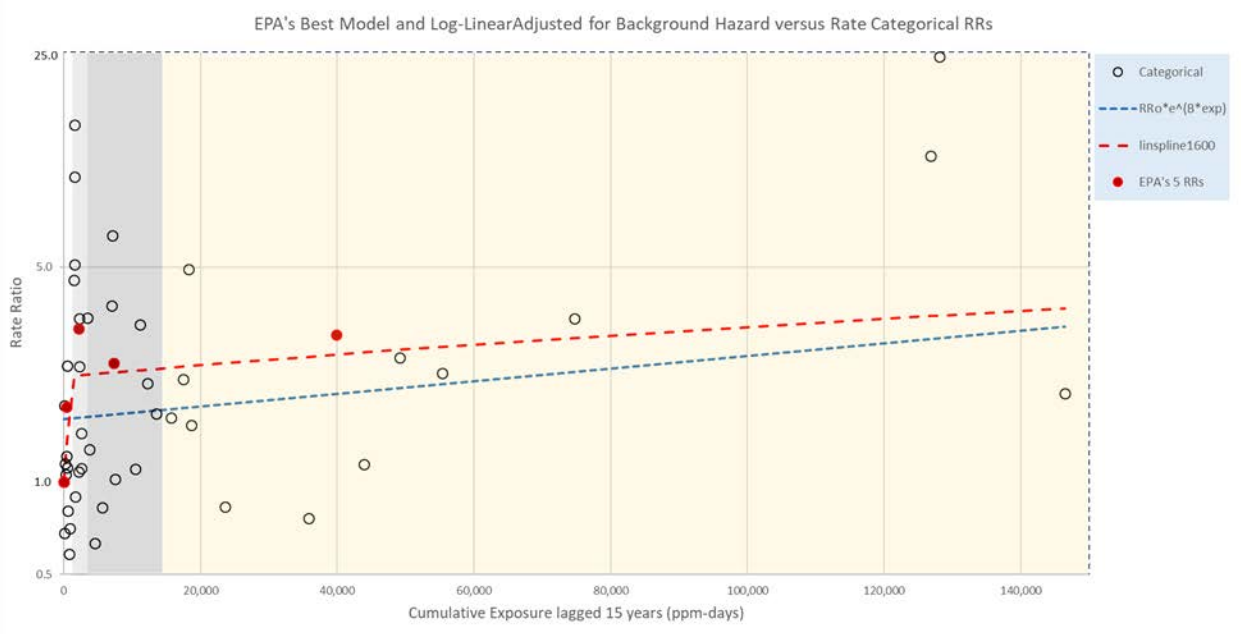


Figure 10. Lymphoid Cancer Death Categorical RRs and the Cox Proportional Hazards and Two-Piece Spline ("knot" at 1,600 ppm × days) Fitted Models for 15-Year Lagged Occupational Doses ≤150,000 ppm × days (NIOSH cohort)

TCEQ's log-linear model in Figure 10 (Figure 22 of the DSD) is plotted *after* adjusting its intercept in the y-axis to account for the different implicit estimated baseline risks (as noted by USEPA) of the non-parametric rate ratios and the log-linear model. The ratio of different implicit estimated baseline risks was estimated using the intercept of the purple-dotted fitted model to the individual estimates of the rate ratios shown in Figure 9 (Figure 20 of the DSD). The purpose of Figure 10 (Figure 22 of the DSD) is to show a fairer visual comparison of the fits of the TCEQ log-linear model, the USEPA two-piece linear spline model (USEPA's selected model for risk evaluation), and the individual estimated rate ratios. This visual comparison in this figure is an attempt to correct the misunderstanding that the RRs estimated by the log-linear model (TCEQ's selected model) are out of line with the RRs estimated non-parametrically (the open circles in the figures). Misinterpretation in the comparison of parametric and categorical (non-parametric) rate ratios used to judge model fit has been published in the peer-review literature (see Valdez-Flores and Sielken 2013).

See the response to Comment #2 regarding the slope not being calculated relative to the lowest exposure category, but rather to the full risk set.

Comment 4:

TCEQ's discussion of the fit of their model to quintiles 2-5 is very misleading because it does not mention the lack of fit of the model against the mortality expected in the unexposed population since the unexposed population has been ignored in the evaluation.

Response:

Because of the nature of the rate ratio going through 1 at zero cumulative exposures, all models result in the same background risk at zero cumulative exposures. That is, the USEPA model and the TCEQ model estimate the same background risk, by definition. This model characteristic is captured in a footnote to Table 32 in the TCEQ DSD [*emphasis added*], "*Quintile 1 is the control (unexposed lagged-out) group with 9 lymphoid cancer mortalities observed and 11.5 mortalities predicted *by all models* with a 95% confidence interval of (6.0, 25.2), which includes the observed 9 lymphoid cancer deaths." Consequently, "the mortality expected in the unexposed population" is the same for USEPA model and TCEQ model, and for that matter, for any relative risk model. The reason the first quintile (the unexposed, lagged-out person-years) was not explicitly included in the comparison in Table 13 is that all models (including USEPA's) result in the same 95% confidence interval that includes the number of observed lymphoid deaths in the NIOSH study.

Comment 5:

Steenland et al. fit the Cox model to their data using the unexposed group as the referent population and did the same thing using log (dose+1) and present their results in Table 7 of their paper. They concluded the log(dose+1) model fit best with the 15-year lag. That model is shown in Figure above (it is the brown solid line). What is noted is the rapid climb from non-exposed to exposed that has been eliminated in the modeling used by the TCEQ because of the intentional decision to rely only on the modeling within the exposed groups. The EPA model (solid red line) accounts for the early rise in risk ratios then flattens out in the higher doses.

TCEQ claims that there is no discernable pattern to the data based on risk ratios where the grouping only includes a single death in each group. What TCEQ has not shown is that there is no confidence in these numbers and that the confidence bounds around the individual risk ratios will be overwhelming. The grouping done by Steenland et al. (red dots, effectively) is the standard epidemiological approach to dealing with these types of data (have enough deaths in an exposure interval to insure reasonable estimates of risk ratios).

Response:

The misconception that TCEQ modeling relied only on "modeling within the exposed groups" has already been addressed above. See the response to Comment #2 regarding TCEQ's modeling including the full risk set, not just the exposed individuals. While EDF is correct that the model that resulted in a parameter estimate most significantly different from zero is the model with the logarithm of cumulative exposure, it must be noted that USEPA rejected this model because of its implausibility. This is indicated on pages H-26 and H-34 of USEPA's 2106 risk assessment [*emphasis added*]:

> "The Cox regression models with log cumulative exposure provided reasonable fits to the data, as described by Steenland et al. (2004) and in the 2006 draft assessment. However, the EPA concluded that these models represented exposure-response

relationships that were *excessively sensitive* to changes in exposure level in the low-dose region, and thus, were *not biologically realistic*."

The TCEQ has shown (Appendix 2 in the revised DSD) that even USEPA's unconventional two-piece linear spline model, which was used as a more realistic alternative to the log cumulative exposure model, is unrealistic in that it significantly over-predicts the number of observed cancer deaths in the NIOSH study.

Next, although the individual data were used for modeling (consistent with USEPA's SAB), as an issue related to USEPA's 2016 assessment, the variability in the estimated quintile rate ratios (RRs) cannot be denied. The purpose of graphing the individual RRs is to show the variability in the underlying individual data, and to help in visually judging model fit and any perceivable shape for an exposure-response. Summarizing the RRs by using fewer grouped individual cases without any confidence intervals masks the true variability in the underlying estimates of the RRs. [Note that the true variability is, however, accounted for by TCEQ correctly calculating p-values and AIC values that indicate USEPA's unconventional model does not fit the data statistically significantly better than the standard Cox proportional hazards model.] Furthermore, the table discussed in the next section shows that the estimated RRs for the individual lymphoid decedents and for the quintiles are consistent as the vast majority of the individual RRs are inside the 95% CIs for the quintile RRs. In the end, USEPA models and TCEQ models fit the same individual-level data, and the variability of the non-parametric estimates are meant to be a visual aid (that is turning out to be misunderstood by some readers).

USEPA used the RRs for quintiles of lymphoid decedents in the NIOSH study to judge model fit to the NIOSH data. Each exposure quintile includes 11 lymphoid decedents and the unexposed (lagged out quintile) includes 9 lymphoid decedents. Table 4-2 in USEPA (2016) lists the estimates of the RRs (ratios of the hazard rate for each exposure quintiles to the hazard rate for the unexposed workers). These RRs are shown in the proposed DSD as purple dots in Figure 19 and red dots in Figures 20 to 22. The RRs for the quintiles are summary estimates of the estimated individual RRs shown by circles in Figures 20 to 22 of the proposed DSD. The RRs for the quintiles are approximately located in the center of the 11 individual RRs included in the quintile and represented by the circles in Figures 20 to 22. The table below shows USEPA's quintile RRs (USEPA calls them ORs) with its corresponding 95% CIs along with the average RR of the 11 individual RRs and the range of the individual RRs (individual RRs are shown as circles in Figure 22). Each of the four exposure RRs in USEPA's 2016 risk assessment (red dots in Figure 22) is the summary of the closest 11 individual RRs (open circles in Figure 22).

Table 4. USEPA Quintile-Specific RRs, 95% Confidence Intervals, and Individual RRs within Each Quintile

Quintile	USEPA's RRs[1] (95% Confidence Interval)	Average of 11[2] Individual RRs in the quintile	Individual RRs in the quintile[3]
2	1.75 (0.59, 5.25)	1.46	0.58, 0.68, 0.71, 0.80, 1.06, 1.11, 1.15, 1.22, 1.77, 2.38, 4.55
3	3.15 (1.04, 9.49)	4.04	0.89, 1.08, 1.11, 1.28, 1.44, 2.38, 3.41, 3.42, 5.11, 9.82, 14.49
4	2.44 (0.80, 7.50)	2.22	0.63, 0.82, 1.02, 1.10, 1.62, 1.67, 2.10, 2.16, 3.25, 3.75, 6.34
5	3.00 (1.02, 8.45)	4.99	0.76, 0.83, 1.14, 1.53, 1.94, 2.26, 2.54, 3.40, 4.93, 11.50, 24.11

[1] Source: Table 4-2 of USEPA's (2016) risk assessment report.
[2] The average of the 11 individual RRs are not statistically significantly different than the quintile RRs estimated by USEPA.
[3] Most individual rate ratios are inside the 95% confidence interval of the RR corresponding to the quintile.

Figure 22 of the DSD (Figure 10 above) and the table show that the alleged steep increase at low cumulative exposures and plateauing of the RRs at higher cumulative exposures is an artifact of summarizing the RRs into quintiles (red dots in Figure 22). The 95% CIs of the quintile RRs and the individual RRs based on each lymphoid decedent shown in the table represent the variability in the NIOSH data for lymphoid cancer. The alleged supra-linearity (steep increase for low cumulative exposures and plateauing at higher cumulative exposures concluded from the red dots in Figure 22) is not supported by the individual RRs (open circles) in Figure 22 (Figure 10 above), which form no discernable dose-response pattern. This figure shows that the two models fit the individual RRs about the same. This is corroborated by the p-values and AICs in Table 38 of the proposed DSD where the linear and the standard Cox proportional hazards model have preferable (i.e., lower) AIC values once the correct degrees of freedom are used for USEPA's selected model. These standard statistical measures of model fit (i.e., p-values, AICs) are calculated so that visual fit need not be relied upon, although visual examination of the actual individual data is consistent with the correctly calculated p-values and AIC values that indicate USEPA's overall supra-linear two-piece spline model does not fit the data better than the TCEQ's standard Cox proportional hazards model.

Figure 22 of the DSD (Figure 10 above) shows USEPA's selected model and TCEQ's selected model plotted against the individual RRs and the quintile RRs for lymphoid mortality. The TCEQ selected RR model goes through 1 for zero cumulative exposures (as do all RR models in Figures 20 and 21 of the DSD). Figure 22, however, shows an adjustment for TCEQ's selected model to account for the difference in the estimated baseline hazard rate estimated by the non-parametric estimates for the quintile RRs and the estimated baseline hazard rate for the

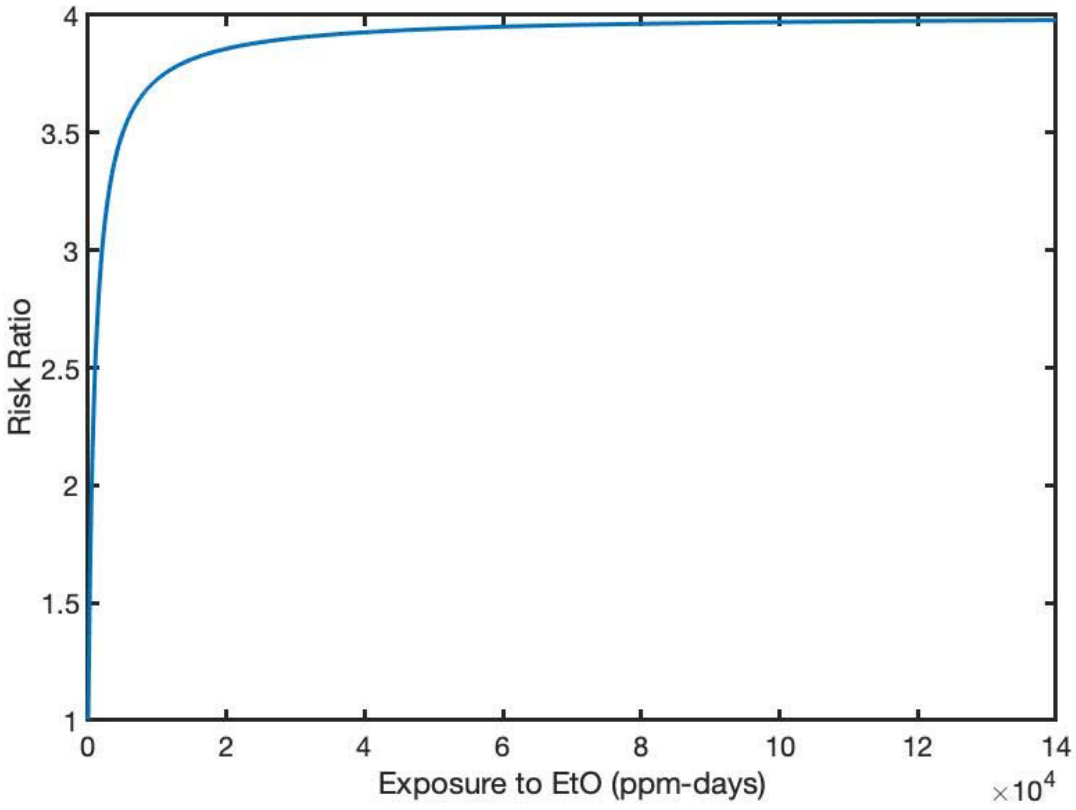
standard Cox model. This adjustment is for the sake of facilitating visual comparison of the two models and the nonparametric rate ratios as indicated by USEPA in the footnote to Figure 4-3 (Figure 19 of the DSD). In this footnote, USEPA states "(Note that, with the exception of the categorical results 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.)" While Figure 22 (Figure 10 above) is explicitly for this stated purpose, again, TCEQ's selected RR model as well as all RR models go through 1 at zero cumulative (see Figures 20 and 21 of the DSD).

In regard to EDF's comments on TCEQ dose-response modeling, most comments provided by the EDF seem to: (1) concern model fit to the observed data; and (2) be based on a belief or misunderstanding that the categorical RRs (non-parametric estimates) are the observed data. However, categorical RRs are not the observed data or the best basis to judge model fit. While USEPA (2016) has footnoted Figure 4-3, the figure remains visually misleading for comparing model fit (for reasons cited in the footnote) yet "visual fit" of parametric models to categorical RRs is still used as an argument in selecting/discarding parametric models (e.g., Table 4-6 of USEPA 2016, EDF comments above). *TCEQ uses a better approach to judge model fit to the observed data*. The number of cause-specific deaths for cumulative exposure ranges are observed as opposed to categorical RRs, which are not observed but rather are estimated. In Table 32 from Appendix 2 of the revised DSD, *the observed number of lymphoid deaths in each quintile were compared with the number of model-estimated lymphoid deaths in the NIOSH cohort. TCEQ's log-linear model estimated the number lymphoid deaths in each of the five quintiles within a 95% CI*. By contrast: (1) *USEPA's selected model (MLE of the two-spline linear model) statistically significantly over-estimated* the number lymphoid deaths in all but one of the four quintiles with non-zero exposure at the 5% significance level; and (2) *USEPA's selected model used for estimating risk (the upper 95% upper confidence limit on the two-spline linear model) statistically significantly over-estimated* the number of lymphoid deaths in all non-zero exposure quintiles. USEPA's model also statistically significantly over-estimated the number of lymphoid cancer deaths for the cohort as a whole. In contrast to USEPA's model, *TCEQ's model neither over- or under-estimated but rather was reasonably accurate*. Additionally, the TCEQ correctly calculates p-values and AIC values to aid in model comparison (Appendix 4 of the revised DSD), which in addition to the carcinogenic MOA do not support adoption of USEPA's overall supra-linear model over TCEQ's standard Cox proportional hazards model.

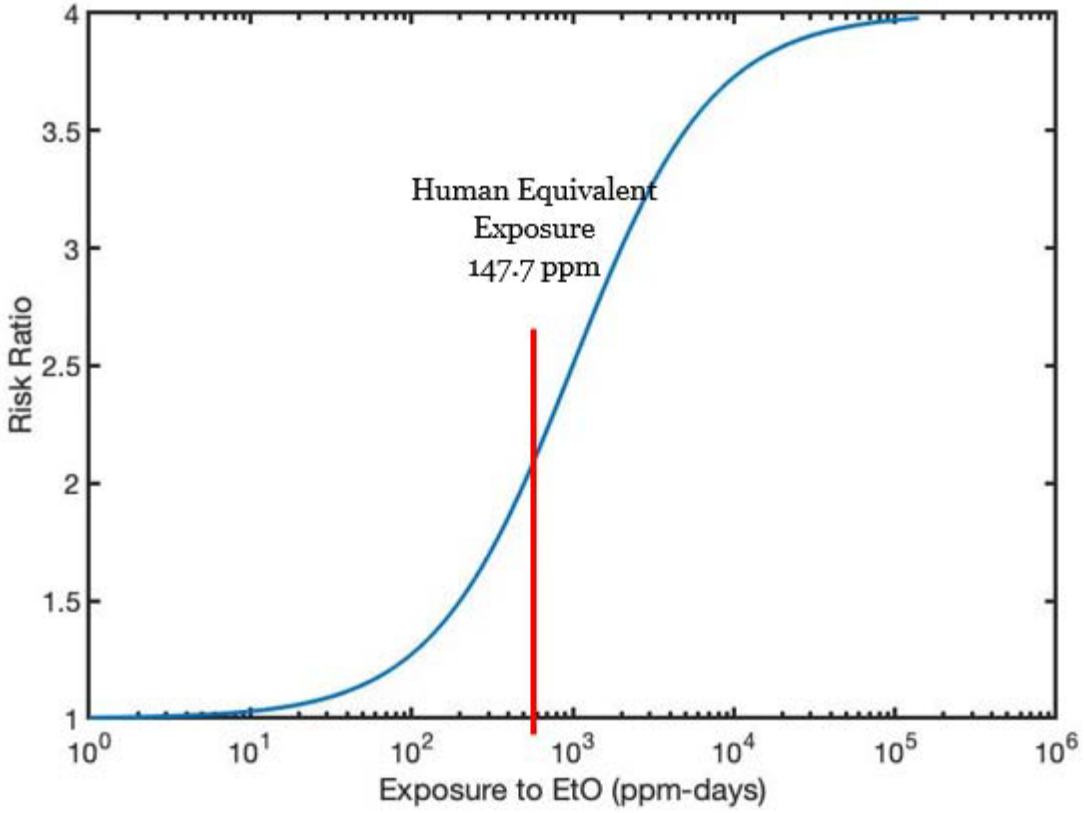
Although model selection for cancer risk assessment should be ideally driven by biology, physiology, and MOA, statistical arguments are oftentimes used for model selection. Statistical arguments need to be clearly understood to make a valid judgement in model selection. Visual representation of summary statistics can be misleading when the summary statistics are believed to be observations. Sound science-based decision making is based on scientific results and not on visual judgement. Though visual aids can be used for communicating results, the visual aids need to be clearly understood by the scientist and all stakeholders. The TCEQ has tried to highlight and clarify these issues.

Comment 6:

Theoretically, the pattern noted here is not unexpected if the chemical investigated is additive to a process that already exists in the body. As an illustration, consider the usual model for Michaelis-Menten kinetics used for most enzymatic reactions in the body (adding 1 to make it range between 1 and 4). To match this to these data, we will use Vmax=4 (the maximum risk ratio), a KD of 1000 (about 8 times higher than the 147 ppm-days mean TCEQ is using for human equivalent air exposure for endogenous EtO) and plot this as a function of EtO ppm-days. After doing this, you get the figure below.



The graph illustrates a rapidly climbing curve reaching its maximum fairly quickly. Putting this on a log(dose) scale allows greater detail of the low-exposure region and provides the following graph:



Here, it is clear that starting from the presumed human background exposure of 147 ppm-days used by TCEQ, there is no curvature at all, but instead the rapidly climbing relative risk that would be expected. The risk-ratio data from Steenland et al. (2004) demonstrates a similar, rapidly-rising risk for low exposures. TCEQ has ignored the fact that additivity to background is likely to lead to linear or supra-linear response.

Response:

The comments above were provided as a hypothetical example, as the included dose-response curves are not plotted using actual data. The second figure shows that in theory, 147.7 ppm-days could fall on the middle linear portion of the dose-response curve with the greatest slope. It is based on the previous figure that shows a supra-linear dose-response based on a model for enzymatic reactions. EDF indicated that the curve would plateau as metabolism becomes saturated (e.g., for a metabolically-activated carcinogen). However, EtO is a direct-acting mutagen and thus the applicability of Michaelis-Menten kinetics for enzymatic reactions as a surrogate for the expected EtO relationship with cancer mortality is very unclear. Even if this class of kinetics were the driver for the relationship of EtO with cancer mortality (not the case for direct-acting mutagens), there would be no data to show where on the curve the

background and occupational EtO concentrations would lie. The commenter suggests that a supra-linear dose-response may be expected if the chemical is additive to a process that already exists in the body. The carcinogenic MOA for EtO, mutagenicity, is one such example. Since lymphoid cancer drove the USEPA carcinogenic assessment, perhaps the most relevant mutagenicity data discussed by USEPA (2016) was that in the bone marrow of mice exposed to EtO by inhalation *in vivo* (Recio et al. 2004), which USEPA indicates is consistent with a *linear* dose-response (C-17 of USEPA 2016). This is inconsistent with the comment's suggestion that supra-linearity should be expected. Even at 200 ppm, much higher than the measured and estimated worker exposure means of 3.5-4.6 ppm (Tables IV and V of Hornung et al. 1994), the mutagenicity dose-response remains linear and has not plateaued (Figure 2 above). These bone marrow mutagenicity data are inconsistent with supra-linearity but consistent with a linear dose-response, despite additivity to a background process producing endogenous EtO already in effect at 0 ppm EtO in air.

Comment 7:

TCEQ argues that the amount that EPA is suggesting as protective of human health is far below the normal human range. This argument does not address the issue of additional human risk, but simply magnitude and associated error of their presumed human background exposure. The argument that the one-standard error estimate away from the estimated human background is much larger than EPA's proposal is confusing statistical noise with the effects in a population. That noise is a function of the response from the people included in the biomonitoring work (people who can clearly have some exposure to EtO, cigarette smoke, ethylene, and other agents in this metabolic pathway), instrument accuracy in calculating values in a urine sample, and the accuracy of the method used to back-calculate to the equivalent air exposure. It also does not mean that the concentrations seen in the population aren't already causing lymphatic cancers because it does not address that question. Additivity to background, as illustrated by the Michaelis-Menten kinetics example illustrated above, demonstrates how rapidly risk can change in the low-dose region for these reactions.

Response:

The normal human range of endogenous or background exposure is primarily provided for context in the DSD. Various examples provide this context, such as USEPA's maximum acceptable air concentration corresponding to an internal dose about 1/40th of the 1st percentile of normal endogenous levels. It is highly biologically implausible that such a small contribution to normal endogenous levels, all of which would be subject to operation of the same protective mechanisms (whether perfect or not), would produce a total internal dose that would be biologically distinguishable from background, giving rise to a biological perturbation resulting in unacceptable risk. This improbability is supported by the TCEQ's reality checks of the USEPA EtO URF on both lymphoid cancer in the key NIOSH cohort and in the U.S. population, as well as in the smoking population specifically.

Regardless of the endogenous or background EtO exposures, the primary deterministic factor for the rejection of USEPA's model under TCEQ guidelines is the USEPA-acknowledged lack of MOA data for EtO supporting an overall supra-linear dose-response curve. The TCEQ does not

state that concentrations seen in the population aren't already causing lymphatic cancers. In fact, the TCEQ assumes this and performs a U.S. population background reality check on USEPA's URF.

The response to Comment #6 addresses the question of Michaelis-Menten kinetics.

Comments from a citizen in the Austin area

Comment 1:

The most alarming thing about the proposed threshold change is that Texas already produces half of the country's ethylene oxide, and the rule change would no doubt lead to increased emissions in a time we are also experiencing hotter years. Ethylene oxide can be explosive when exposed to air and heated, so you can understand why a resident of an urban center would be concerned about an increase in toxic gas in the air.

Response:

Although EtO is a flammable gas, it takes very high concentrations in order for it to be explosive. The lower explosive limit, or LEL, for EtO is 3.6% volume in air, which is equivalent to 36,000 ppm. This concentration is significantly (i.e., orders of magnitude) higher than could be expected in ambient air (e.g., compared to the TCEQs revised EtO chronic ESL of 2.4 ppb).

Comment 2:

The assessment that is cited in the paper on the commission's website is based on comparisons to 'endogenous values', the calculation of and data sources for which are never discussed. The value of 'endogenous' EtO is not at all a consensus value. The article by Kirman and Hays cited as the basis of the assessment throughout the document was commissioned by the industry trade group, the American Chemical Council, as cited in the conflicts of interest section toward the end of the article "The analysis presented here was funded by the Ethylene Oxide Panel of the American Chemistry Council (contract 5478)."

The 'endogenous value' holding up the analysis has not been independently verified. The interests of the American Chemical Council, driven by nebulous 'market forces', and the interests of the general public, with which the TCEQ is entrusted, diverge in obvious ways. The American Chemical Council is entrusted with upholding the economic interests of its members, who are driven by the maximization of profit regardless of environmental impact. This is a story we have read many times.

Response:

Across a given field of study, scientists rarely come to a consensus on single values representing anything. In this case for example, a survey across regulatory and other agencies of health-protective environmental media concentrations (e.g., air, soil, water, fish tissue) for well-studied chemicals would yield many disparate values with no consensus value ever emerging. Kirman and Hays (2017) is a peer-reviewed article appearing in the scientific literature. Science is paid for and conducted by those who have interest in better understanding the underlying science and important scientific issues, including government, industry, academia, or others. Although TCEQ understands the commenter's concerns (i.e., ACC's interests), it should be noted that: (1) Scientific journals recruit independent and objective peer review experts to verify the scientific soundness of article approaches; (2) the TCEQ used a systematic review process that included risk of bias analysis for all articles, including the ones referenced that were published

as peer-reviewed scientific articles (fully described in Appendix 1 of the DSD); and (3) while TCEQ used Kirman and Hays (2017) for context, it played no deterministic role in choice of the modeling approach or derivation of the URF.

Comment 3:

The endogenous equivalent value is a calculation of the ambient ethylene oxide air concentration to which non-smoking test subjects were exposed in the months leading up to sampling. The value is a calculation of the baseline exposure of these test subjects and not the endogenous production of ethylene oxide, the two being distinct scientific concepts. The 'endogenous values' based on reporting by Kirman and Hays reflect the baseline exposure of the study participants which has nothing to do with endogenous production at all. The value is an extrapolation of ambient EtO concentrations based on hemoglobin adduct quantitation, which does not discriminate between endogenous and exogenous exposure. To call this value 'endogenous' is to be factually incorrect and misleading to the public for the sake of profit at the expense of our health and environment.

Response:

The TCEQ appreciates the commenters note about the distinction between typical environmental exposure and endogenously-produced EtO contributions to EtO-hemoglobin adduct formation. For EtO, endogenous production overwhelmingly dominates the distribution. Based on the linear relationship between EtO exposure concentrations and the HEV biomarker of exposure, typical environmental levels of EtO would not significantly add to endogenous levels. For example, USEPA's maximum acceptable air concentration corresponds to an internal dose only about 1/40th of the 1st percentile of distribution in Kirman and Hays (2017). Subtracting the very small contribution from typical environmental levels would have no practical effect on the distribution, which is dominated by endogenous levels. The TCEQ also points out that the workers in the NIOSH cohort also had typical environmental exposure outside the workplace as a small part of their background exposure (i.e., endogenous plus a very minor contribution from environmental). As the dose-response assessment is for occupational exposures over and above such a baseline, the comparisons made by TCEQ are all the more appropriate. However, subtracting the insignificant contribution from environmental exposure would make no difference in terms of the outcome of comparisons to USEPA-risk based values or the doses associated with carcinogenicity in epidemiology studies. The reasonableness of endogenous levels cited in Kirman and Hays (2017) and the high likelihood that environmental EtO levels had no significant effect on the data is further strongly supported by endogenous data in laboratory control animals. More specifically, the TCEQ notes that the reported mean human background endogenous HEV level of 21.1 pmol/g Hb appears very reasonable given background HEV levels in control rats (≈42-50 pmol/g Hb) and mice (≈58-100 pmol/g Hb) (Walker et al. 1993, 2000). This information was added to the DSD for clarification.

Comment 4:

The ethylene oxide assessment is also difficult to read because data are frequently presented in relation to the non-consensus 'endogenous values' discussed above that are the linchpin of the

assessment. To compare real values and examine the relationships the authors put forth, a reader must pull all the relevant originals and comb through them for the data to recreate the basis for the likely inaccurate conclusions in the paper. An average Texan who wants to be engaged in the processes of her government and have a say should not have to work so hard to understand what is scientifically proven and what is vapor. The authors are obfuscating their wrong assertions by massaging numbers and conflating concepts.

Response:

Endogenous data are used for context and are not the linchpin of the assessment, which is essentially the lack of EtO MOA data supporting the adoption of a non-standard, overall supra-linear dose-response model. The TCEQ endeavors to present all calculations and methods in the DSD in a straightforward and repeatable manner, and has tried to clarify our discussion of endogenous levels of EtO in this DSD.

Comment 5:

The evaluation is also based on mortality from the cancers examined. Mortality rates vary as medical advances reduce mortality compared to incidence. Therefore, basing a calculation on mortality is muddying the waters. Incidence, however, directly correlates exposure with human impact. Cancer treatment itself is a significant medical expense and personal trauma that should not be discounted. Furthermore, the time frames of the studies make a mortality comparison questionable because the relationship between incidence and mortality was much closer in the older Union Carbide study than the more recent NIOSH study. NIOSH would be expected to have lower mortality rates due to medical progress over time in improving the survival rate of cancer incidence. In short, mortality is a measure of medicine and incidence is the true measure of toxicity.

Response:

The TCEQ recognizes the terrible personal and societal toll of cancer, which touches the lives of every family at one time or another. In the interest of public health, the agency is committed to the accurate prediction of the risks that environmental chemicals pose so that effort and resources can be appropriately prioritized and committed to mitigating those exposures commensurate with the risks they most realistically represent. Most often, epidemiology studies must rely on mortality data as opposed to incidence data, even though incidence information is a much better marker of potential toxicity, because there are no incidence data available. This is the case for lymphoid cancer data for the key NIOSH cohort, and is a primary reason why the TCEQ has chosen to use the 95% UCL on the slope as the more conservative estimate, in addition to incorporating ADAFs in consideration of childhood exposure. These conservative assumptions are health-protective and include protection for unknowns like the relationship between incidence and mortality. Furthermore, the health-protective assumptions made by TCEQ are above and beyond the presumption of carcinogenic effects of exposures to EtO. Workers in the NIOSH and UCC studies (and several other studies) do not show a statistically significantly increasing cancer response relationship with increasing exposure to EtO.

Comment 6:

Finally, the assumptions and key evaluations chosen by the authors betray the purpose of re-evaluating the allowable limits in the draft assessment provided by the TCEQ, which appears to be to justify increased emissions by businesses located in Texas. In contrast, the US EPA IRIS report for ethylene oxide released in 2016 establishes throughout the document that its purpose is to develop a protective risk assessment that takes into account not only worker safety, but sufficient margins of protection for the general public and children.

Response:

The TCEQ was compelled by scientific duty to conduct our own EtO dose-response modeling when upon reviewing the USEPA assessment in detail, significant scientific flaws became apparent. These are outlined in the DSD and included such things as incorrectly calculated p-values and AIC values which were erroneously used by USEPA to justify an unconventional, overall supra-linear model that TCEQ has shown to be over-predictive both for the key NIOSH cohort and the U.S. population. The TCEQ DSD uses the 95% UCL on the slope in consideration of public health protection, as well as ADAFs in consideration of childhood exposure.

Comments from the Ethylene Oxide Sterilization Association

Comment 1:

The URF for EO based on lymphoid cancer derived by TCEQ is 2.5E-6 per parts per billion (ppb) (1.4E-06 per µg/m^3) and results in a risk-based air concentration of 4 ppb at the no significant excess risk level of 1 in 100,000. This value is much more realistic and defensible than the EPA IRIS URF of 9.1E-3 per ppb (5.03E-3 per µg/m^3), which yields a 1 in 100,000 risk concentration of 1 part per trillion (ppt). EOSA agrees with TCEQ that the EPA IRIS assessment is not adequately supported by current scientific data. The EPA IRIS assessment should not be used as the basis for the exclusion of other estimates of cancer risk that are based on other robust studies and modeling methodologies with more representative exposure estimates. Reliance on the flawed, overly conservative, and outdated EPA IRIS assessment would result in disastrous consequences to the healthcare industry and public health in the United States. Decisions on how best to protect public health cannot be made on such demonstrably flawed science.

Response:

The TCEQ agrees and appreciates this comment.

Comment 2:

Though EOSA supports TCEQ's statements regarding the inaccuracies of EPA's modeling approach, the non-threshold effects screening level (ESL) of 4 ppb is overly conservative when you consider endogenous and background ambient EO levels. EPA's December 2016 IRIS assessment relies exclusively on a National Institute of Occupational Safety and Health (NIOSH) epidemiology study of sterilizer workers. The NIOSH study used a model to estimate exposures prior to 1978 because there were virtually no measured EO concentration data for sterilization workers prior to 1978. The NIOSH exposure model estimates job exposures that were lower than levels observed from 1978 and later. This would be an unusual pattern of historical exposure, based on industry experience with sterilization operations. This unusual pattern was also noted by the EPA Science Advisory Board (SAB) during its review of the IRIS assessment.

Response:

The TCEQ agrees that, as stated in the DSD, there appears to be appreciable uncertainty stemming from the lack of EtO exposure data for the NIOSH cohort prior to the time air monitoring data collection began. These exposures for much of the cohort would have been relatively high and significantly contributed to cumulative exposure estimates (ppm-days, both unlagged and lagged), which appear likely to be biased low (e.g., Bogen et al. 2019, Li et al. 2019). The USEPA SAB agreed that these exposure estimates are likely of lower reliability (because there were no exposure measurement data that could be included in the exposure model prior to 1979) and actual EtO exposures were likely to have been higher than reflected in the estimates (p. I-41 of USEPA 2016). Use of likely low-biased exposure estimates for earlier periods of time lend additional conservatism to the assessment. However, the TCEQ disagrees that in the absence of additional analyses, the TCEQ assessment can be characterized as "overly" conservative (i.e., that the proposed long-term ESL is significantly lower than it would

have been had the earlier monitoring data been available). Similarly, data on endogenous levels and ambient levels do not lead or compel the TCEQ to conclude that the assessment is overly conservative. For example, the TCEQ URF for lymphoid cancer passes the NIOSH worker and U.S. population reality checks.

Comments from the Houston Health Department

Comment 1:

The Houston Health Department opposes the approval of the Texas Commission on Environmental Quality Developmental Support Document for Ethylene Oxide which conflicts with and presents a lower toxicity value for ethylene oxide than that used by the Environmental Protection Agency (EPA). According to the 2014 EPA National Air Toxics Assessment (NATA), ethylene oxide is of grave concern for Houstonians, not only is ethylene oxide the cancer risk driver for the highest risk census tracts in Houston/Harris County, census tract 343100 has the 7th highest cancer risk from ethylene oxide and the 11th highest total cancer risk in the nation (Table 1, Figure 1). In fact, ten census tracts in Harris County are in the top one hundred highest ethylene oxide cancer risks in the nation, out of a total of 76,727 census tracts (with another census tract just outside the top one hundred). Lowering the toxicity value will remove pressure from understanding and mitigating exposure to this pollutant.

Response:

Although the NATA data can be informative, it has a number of limitations that must be kept in mind when interpreting the results. Multiple assumptions go into the calculations that create the NATA data, which means that the output from the NATA cannot be specific for individuals or small areas. As stated on the USEPA website, "NATA results are best applied to larger areas – counties, states and the nation. Results for smaller areas, such as a census tract, are best used to guide follow-up local studies." Additionally, one of the factors used to calculate cancer risk within the NATA is the USEPA URF. The USEPA URF for EtO was updated in 2016, and this URF was used in the most recent 2014 NATA. This explains why the 2011 NATA data does not show the same EtO cancer risk. The TCEQ has determined that the dose-response model used by the USEPA to develop their URF is not supported by the scientific data for EtO, and therefore did not adopt it, and would therefore not agree with the putative excess risk presented in the NATA.

At the county level, while Harris County has relatively high reported National Emissions Inventory (NEI) EtO emissions per square mile (i.e., 6.6E-03 tons/square mile is ≈188 times higher than the US at 3.5E-05 tons/square mile), the incidences of leukemia (13.0 (95% CI of 12.5, 13.5)), non-Hodgkin's lymphoma (16.9 (95% CI of 16.3, 17.5)), breast cancer (111.9.4 (95% CI of 109.9, 114.0)), as well as all cancers combined (400.1 (95% CI of 397.2, 403.1)) are all statistically significantly lower than in the general US population.

Comment 2:

In general, the Houston Health Department advocates for consistency of toxicity values with the EPA. In this case, we are especially concerned with the departure from the use of the EPA toxicity values because the proposed value is less protective, and many residents are potentially affected. There is a high degree of uncertainty in understanding the toxicity of individual chemicals on humans, but even more so in Houston, where residents are exposed to mixtures

of other toxic chemicals. The health of Houstonians is better protected with the current EPA toxicity value.

Response:

Across a given field of study, scientists rarely come to a consensus on any toxicity value. A survey across regulatory and other agencies of health-protective environmental media concentrations (e.g., air, soil, water, fish tissue) for well-studied chemicals would yield many disparate values with no “consensus value” ever emerging. The TCEQ has determined that the EPA statistically significantly over-estimated the risk associated with ethylene oxide exposure, and therefore we developed our own scientifically-sound assessment.

Comments from Harris County, TX

Comment 1:

The EPA Unit Risk Factor (URF) model for ethylene oxide, referenced in the DRA [Dose Response Assessment], is clearly a more conservative approach to determine a health based screening value than the method proposed by TCEQ. The TCEQ Cox Proportional Hazard method uses a very different approach and results in a level that is higher and may not be protective of human health. The proposed TCEQ health based screening value is 4,000 times less protective than the EPA value. When human health is at risk, and sensitive populations are at risk of exposure, increasing a health-based value by over three orders of magnitude must be supported by ample scientific data. Harris County requests that TCEQ coordinate with the EPA and reach a concurrence prior to the determination of a final value.

Response:

Best science is not determined by the method that results in the lowest value. The issue in this case is which model assessment most accurately predicts cancer risk. The TCEQ has demonstrated that USEPA's selected model assessment over-predicts lymphoid cancer in the key NIOSH cohort and the U.S. population, its selection was based on incorrectly calculated p-values and AIC values, and it is biologically implausible. By contrast, TCEQ's assessment is supported by all these considerations. Across a given field of study, scientists rarely come to a universally accepted consensus. A survey across regulatory and other agencies of health-protective environmental media concentrations (e.g., air, soil, water, fish tissue) for well-studied chemicals would yield many disparate values. The TCEQ is always willing and ready to discuss our assessments with the USEPA or any other group, and is willing to reach consensus on scientific issues if such consensus can be reached.

Comment 2:

Section 3.4.2 of the DRA notes that there is a lack of chemical specific data on susceptibility from early-life exposures and as a result, the TCEQ utilized default ADAF numbers.

Response:

There is a general lack of chemical-specific data for most chemicals on susceptibility from early-life exposures, including for EtO. For this reason, both USEPA and TCEQ use ADAFs in consideration of childhood exposure when a chemical has been identified as acting through a mutagenic MOA, consistent with USEPA guidance. The TCEQ DSD now calculates the ADAF-adjusted ESL consistent with equation 5-17 of the TCEQ guidelines (TCEQ 2015), rounding it to two significant figures at 2.4 ppb.

Comment 3:

Harris County urges the TCEQ to conduct a study to determine the actual background ethylene oxide concentrations in Harris County and impacts of ethylene oxide on children and pregnant

women. This data is necessary to understand our residents' chemical burden and health impacts.

Response:

Such ecological studies have significant uncertainties and are typically not useful in determining causation (i.e., attributing effects to a given chemical exposure) or for regulatory dose-response assessment. Previous studies at various locations have yielded mixed results and are of little regulatory use (see TCEQ response to Comment #7 from the Union of Concerned Scientists).

Comment 4:

Harris County requests that the TCEQ hold a public meeting to explain to the public the scientific process utilized in this determination as well as to present supporting data.

Response:

The process is outlined in the TCEQ toxicity factor guidelines (https://www.tceq.texas.gov/assets/public/comm_exec/pubs/rg/rg-442.pdf), and the DSD itself presents both key and supporting data. Both have undergone previous public comment.

Comments from the Natural Resources Defense Council

Comment 1:

TCEQ process violates EPA peer review requirements.

EPA's process for developing its EtO IRIS assessment was consistent with its peer review requirements as described in the EPA Peer Review Handbook (4th Edition, 2015), and the Final Information Quality Bulletin for Peer Review (OMB, 2004). In stark contrast, TCEQ's EtO assessment and proposed risk estimate has not been vetted by an external scientific advisory committee or any other appropriate scientific peer review committee, and has not undergone any public peer review, scientific scrutiny, or public comment before this comment period.

The peer review process that TCEQ should undertake should be transparent, include EPA input, and be accountable to the recommendations that arise from that process. The Handbook and Bulletin require documents that are "highly influential," "novel, controversial, or precedent-setting," or have "significant interagency interest" to undergo peer review before being implemented. EPA has done this, while TCEQ has not.

Response:

Although the TCEQ is not required to do a peer review (TCEQ 2015) nor do we fall under the same requirements as the IRIS program, the TCEQ has decided to conduct an independent external scientific peer review in the first quarter of 2020 to strengthen our assessment. The TCEQ has also had the benefit of SAB and public comments provided on the USEPA (2016) assessment, as well as extensive public comments on the DSD from a diverse group of commenters (e.g., NGOs, academia, industry, private citizens).

Comment 2:

TCEQ fails to address all cancer and non-cancer risks.

TCEQ narrowed the focus of its assessment to lymphoid cancers only, disregarding the elevated breast cancer incidence in female workers. This results in an underestimate of the risks posed by EtO, and is one of the most significant differences between the TCEQ assessment and the EPA IRIS assessment, which included breast cancer data.

Response:

As with USEPA's 2016 assessment, the TCEQ DSD was limited to a carcinogenic dose-response assessment. The TCEQ decided to move forward with the cancer risk assessment first in order to expedite the process. Although long-term limits based on cancer risk are expected to determine allowable emissions in both the short- and long-term, the agency plans to return to the non-cancer assessment in the near future. TCEQ does have interim ESLs in place in recognition of these concerns.

The TCEQ evaluated breast cancer as a candidate cancer endpoint. USEPA's assessment, however, is driven by lymphoid cancer as the primary contributor to the URF. Subsequent to USEPA's 2016 assessment and TCEQ's systematic review of the peer-reviewed literature for EtO, recent meta-analyses of available studies have been published (Marsh et al. 2019, Vincent et al. 2019). These new meta-analyses included Steenland et al. (2003, 2004) and the smaller Mikoczy et al. (2011) study cited by USEPA (2016), as well as other studies, and reported breast cancer meta-RRs of 0.97 (0.80, 1.18) (Marsh et al. 2019) and 0.92 (0.84, 1.02) (Vincent et al. 2019). The Marsh et al. study concluded [*emphasis added*], "Evaluations of workers exposed during sterilization processes *do not support the conclusion that EO exposure is associated with an increased risk of breast cancer*." Similarly, the Vincent et al. (2019) study concluded, "Higher quality epidemiological studies demonstrated no increased risk of breast cancers." In addition to evaluating epidemiological evidence, Vincent et al. (2019) evaluated animal study results and concluded that they provide no strong indication that EtO causes mammary tumors. These recent meta-analyses and other information (IARC 2019) further support TCEQ's decision not to base the URF on breast cancer. Human data are by far the most relevant for derivation of human toxicity factors, and the human data themselves are inconclusive (as acknowledged by USEPA 2016). The weight of evidence for EtO-induced breast cancer is now discussed in a new appendix (Appendix 6) in the revised DSD.

That being said, the assessment does not ignore risk to women as the lymphoid cancer estimate is shown to be conservative for women. The TCEQ URF is ultimately based on the same key epidemiological study as USEPA's URF. The study included both male and female workers. The dose-response assessment evaluates lymphoid cancer risk to males and females combined as well as males alone. Analyses with both males and females combined compared to results based on males alone show that use of risk results based on males alone is conservative for estimating lymphoid cancer risk in females, which is demonstrated in the DSD. Thus, a conservative long-term ESL and URF is used to evaluate lymphoid cancer risk to women, resulting in a lower long-term ESL for everyone (compared to using modeling results for males and females combined).

Table 5. Human Studies Relevant to the Breast Cancer Weight of Evidence

Study Type	Workers (n)	EtO Exposure Level (ppm)	Observed (O)	Expected (E) [a]	O/E (95% CI)
Individual Studies					
Steenland et al. (2003)	7,576 female workers	Median ≈14 ppm-years; Mean >1 ppm [b]	230 [c]	258.4	**0.89** [d] (0.78, 1.01)
Steenland et al. (2004)	18,235 workers (≈55% female)	Mean of 26.9 ppm-years	103	102	**0.99** (0.84, 1.17)
	only female workers				**0.99** [e] (0.81, 1.20)
Mikoczy et al. (2011)	2,046 workers (≈60% female)	Means ≤1.11 ppm; Peaks up to 40-75 ppm	33	38.54	**0.86** [f] (0.59, 1.20)

Study Type	Workers (n)	EtO Exposure Level (ppm)	Observed (O)	Expected (E) [a]	O/E (95% CI)
	615 female 287 female 295 female	Mean of 0.02 ppm in lowest cumulative exposure group Mean of 0.021 ppm in middle cumulative exposure group Mean of 1.11 ppm in highest cumulative exposure group			**0.52** [g] (0.25-0.96) 1.06 (**0.58**, 1.78) 1.12 (**0.65**, 1.79)
Norman et al. (1995)	928 female	TWA 50-200 ppm; 5-20 ppm post-corrective action 1980	12	7.64	1.57 [h,i] (**0.90**, 2.75)
Coggon et al. (2004)	1,012 female	TWA generally < 5 ppm; Peaks up to > 700 ppm	11	13.1	**0.84** [j] (0.42, 1.50)
Hogstedt et al. (1986)	153 female	TWA 20±10 ppm	0	---	**No breast cancer reported**
Meta-Analysis Studies					
Marsh et al. (2019) [k]					**0.97** (0.80, 1.18)
Vincent et al. (2019) [k]					**0.92** (0.84, 1.02)

[a] Based on external referent US population; see the text for information regarding why a healthy worker effect should not be expected for breast cancer incidence, an endpoint relied upon by USEPA (2016).

[b] Using the 233 cases with interviews as a surrogate, mean exposure level would be expected to be > 1 ppm since the mean is higher than the median in a lognormal distribution, median cumulative exposure for the 233 cases was 14.0 ppm-years, and mean years exposed was 13.0 (Table 2 of the study), so mean cumulative exposure >14 ppm-years/mean duration of 13 years = >1 ppm mean exposure.

[c] From Table 3 of the study based on workers whose exposure did not lag out using a 15-year lag period, consistent with USEPA (2016) and TCEQ; expected (E) value of 258.4 was calculated (i.e., E=O/0.89).

[d] For a 15-year lag, consistent with that used by USEPA (2016) and TCEQ.

[e] Breast cancer did not show any overall excess, although there was an excess in the highest cumulative exposure quartile (>12,322 ppm-days) using a 20-year lag and internal exposure-response analyses found positive trend for breast cancer using the log of cumulative exposure with a 20-year lag but not cumulative exposure (Tables 1, 5, and 8 of study).

[f] From Table 3 of Mikoczy et al. (2011) and includes induction latency period of ≥15 years, consistent with that used by USEPA (2016) and TCEQ.

[g] This statistically significantly decreased breast cancer risk occurred in female workers exposed to a mean of ≈20 ppb EtO; this inordinately decreased SIR for the lowest cumulative exposure group produced statistically increased SIRs for higher cumulative exposure groups which did not experience increased breast cancer risk compared to the general population despite EtO mean exposures up to ≈1,110 ppb and more robust female worker data suggest that it represents an anomalous study artifact.

[h] For the most appropriate method identified by the study authors (Method 2) for the longest follow-up period (through 1987) with the most appropriate/matching SEER rates (through 1987) used to calculate the expected number (E).

[i] Includes two breast cancers diagnosed within 1 month of employment; reasonably excluding these two breast cancers diagnosed within 1 month of beginning work would not be expected to significantly reduce person-years but would result in a lower and still statistically insignificant estimated O/E (e.g., 10/7.64 = 1.31).

[j] For female workers with known continuous workplace exposure, the breast cancer mortality SMR was 0.70 (5 observed vs. 7.2 expected).

[k] This meta-analysis included all the individual studies above except for Hogstedt et al. (1986), which found no breast cancers and therefore did not report any effect estimate for breast cancer.

Comment 3:

TCEQ fails to address environmental racism in cancer death rates.

TCEQ disregarded EtO-associated breast cancer risks in women workers by conducting assessments based only on morbidity, that is, deaths from breast cancer. Because many women that are diagnosed with breast cancer can survive it, this make is seem as if the risk is negligible. In addition to underestimating risks by counting only women that die of breast cancer, TCEQ's approach also underestimates risks to women of color more than white women. This is because cancer survival depends on access to quality health care, routine medical screening procedures, and medical insurance to allow timely and effective treatments – all of which have significant racial bias. That is, although over time there has been a decline in deaths from breast cancer, not all women have benefited equally. This is evidence by the striking divergence in mortality trends between black and white women beginning in the early 1980s. As treatment for breast cancers has improved, the racial disparity widened; in 2015, breast cancer death rates were 39% higher in black than white women. This is particularly relevant for EtO and other contaminating facilities that are disproportionately co-located in areas of Texas such as the Houston Ship Channel communities that are predominantly low-income and communities of color.

Response:

When deriving a chemical toxicity factor, the TCEQ (and the USEPA) incorporate multiple layers of conservatism into the value so that all members of the general public are protected, including sensitive subpopulations. The long-term ESL applies equally for all people regardless of race, ethnicity, socioeconomic status, etc., and assumes 24 hours/day, 365 days/year exposure to be protective of even those most exposed to ambient EtO air concentrations. The health-protective limit established by the TCEQ includes the following conservative considerations:

1) The lack of human data sufficient to establish EtO as a human carcinogen despite occupational exposures up to millions of times ambient levels;
2) The lack of statistically significantly increasing risk with exposure to EtO;
3) Using the endpoint (lymphoid cancer), study cohort (NIOSH), and sub-cohort (males) that predicts the largest estimated risk of seventy-two non-increasing exposure-response models;
4) Uncertainty in the model parameter estimate (i.e., use of the 95% UCL);

5) Assumed *continuous* exposure to EtO, 24 hours a day, 7 days a week, for every week of every year, from birth until 70 years of age; and
6) Adjustment for early-age exposures using ADAFs.

The TCEQ did consider breast cancer incidence in our EtO evaluation for this DSD, but ultimately concluded that evidence showing that EtO causes breast cancer is unconvincing. Please see the response to Comment #2 above and more specifically, Appendix 6 of the revised DSD for a more detailed discussion.

Comment 4:

TCEQ mis-represents endogenous exposures.

Understanding the endogenous exposures is a scientifically critically important component of the EPA's EtO assessment, whereas it is mis-represented by TCEQ in its assessment. Although the body produces EtO endogenously (through cellular metabolic processes), it has some defense mechanisms, albeit imperfect ones, to deal with some level of endogenous exposure. Given that both breast and lymphoid cancers are fairly common, it is possible that some may be due to endogenous EtO levels, suggesting that the body's defense mechanisms may be largely overwhelmed by additional exogenous EtO from preventable industrial sources, especially when considered across the whole population. While EPA has addressed this in its more sophisticated and scientifically accurate assessment, the TCEQ assessment simply 'zeroes out' the cancer risks at lower exposures as if they do not exist, and – even more flawed – that the body's cellular defense mechanisms will make additional cancer risks 'go away'. This makes as little scientific sense as it sounds, and even less when considered across a diverse population that includes sensitive individuals.

Response:

The normal human range of endogenous or background exposure is primarily provided for context. Various examples provide this context, such as USEPA's maximum acceptable air concentration corresponding to an internal dose about 1/40th of the 1st percentile of normal endogenous levels. It is highly biologically implausible that such a small contribution to normal endogenous levels, all of which would be subject to operation of the same protective mechanisms (whether perfect or not), would produce a total internal dose that would be biologically distinguishable from background, giving rise to a biological perturbation resulting in unacceptable risk. This improbability is supported by the TCEQ's reality checks of the USEPA EtO URF on both lymphoid cancer in the key NIOSH cohort and in the U.S. population, as well as in the smoking population specifically.

Regardless of the endogenous or background EtO exposures, the TCEQ uses the lack of MOA data justifying an overall supra-linear dose-response model as the primary determinant for rejecting the unconventional two-piece spline model and using a standard model (Cox proportional hazards model) to estimate additional risk. The TCEQ does not state that concentrations seen in the population aren't already causing lymphatic cancers. In fact, the TCEQ assumes this and performs a U.S. population background reality check on USEPA's URF.

Comment 5:

NRDC supports the calls of Texas residents and others in calling for TCEQ to simply adopt the U.S. EPA scientists' determination and cancer risk factor for ethylene oxide, finalized by the IRIS program in 2016. The EPA IRIS determination reflects the best available science.

Response:

The TCEQ's DSD fully considers the USEPA's derivation of their EtO URF and finds several serious scientific flaws in USEPA's 2016 assessment. The DSD then carefully and completely lays out the scientific basis for our EtO URF derivation, based on the best available science and supported by multiple lines of evidence, some of which are newly available. The revised DSD and accompanying analyses will be undergoing peer review in the first quarter of 2020.

Comments from the Achieving Community Tasks

Comment 1:

We urge the Texas Commission on Environmental Quality (TCEQ) to follow the best available science and not to weaken protections for the thousands of Texans exposed to the carcinogen ethylene oxide. We respectfully request that TCEQ not finalize the proposed Development Support Document (DSD), and instead adopt the robust, final, peer-reviewed cancer risk factor that the Integrated Risk Information System (IRIS) of the U.S. Environmental Protection Agency (EPA) finalized in 2016.

Response:

The TCEQ strives to use the best available science for all our chemical risk assessments. It is our process to review the assessments conducted by other agencies to determine whether or not their methodologies and assumptions are consistent with our guidance (TCEQ 2015), and if not we conduct our own risk assessment. The TCEQ's DSD fully considers the USEPA's derivation of their EtO URF and finds several serious scientific flaws in USEPA's 2016 assessment. The DSD then carefully and completely lays out the scientific basis for our EtO URF derivation, based on the best available science and supported by multiple lines of evidence, some of which are newly available. The revised DSD and accompanying analyses will be undergoing peer review in the first quarter of 2020.

Comment 2:

We ask that TCEQ seek external scientific peer review and provide adequate time for public notice and comment, if it continues to consider the proposed DSD.

Response:

The TCEQ's EtO DSD has undergone a 90-day public comment period, and TCEQ will conduct an external independent expert peer review in the first quarter of 2020.

Comments from the Advanced Medical Technology Association

Comment 1:

The Advanced Medical Technology Association (AdvaMed), the national association of medical technology providers, supports the Texas Environmental Quality Commission's (Commission) reasonable, risk-based approach to the exposure assessment for ethylene oxide and we also support the Commission's rejection of the faulty Integrated Risk Information System (IRIS) approach.

Response:

The TCEQ appreciates your support of this risk assessment.

References

Please see the TCEQ DSD for the referenced cited in this document.